

Ein Virologie-Dokumentationssystem auf einem Kleinrechner

B. Blümmel, U. Eckert, M. Rothmund, G. K. Wolf

Zusammenfassung

Die virologisch-serologische Diagnostik stützt sich auf den Vergleich aktueller Befunde mit Vorbefunden. Die Entwicklung der Software für ein solches umfangreiches Dokumentationssystem sollte sich moderner Entwicklungstechniken, wie der schrittweisen Verfeinerung, der strukturierten Programmierung und der projektbegleitenden Dokumentation bedienen. Diese Methoden erweisen sich als Voraussetzung für ein effizientes kollegiales Teamwork.

Summary

Comparison of actual serological results with previous data is a basic feature in clinical virology. In developing such an information system, we used software engineering techniques such as stepwise refinement, structured programming and immediate project documentation. Those methods proved to be essential for efficient as well as collegial teamwork.

1 Einleitung

Eine der wesentlichen Aufgaben eines Virologischen Institutes ist die Bestimmung von Antikörpertitern in eingesandtem Untersuchungsmaterial von Patienten. Isolierte Antikörpertiter sind nicht sehr aussagefähig, weil erhöhte Titer über lange Zeiträume erhalten bleiben. Deswegen ist ein Rückschluß auf den aktuellen Krankheitszustand nur dann möglich, wenn ein älterer Befund vom selben Patienten vorliegt. Um den Vergleich mit evtl. vorhandenen Vorbefunden durchführen zu können, ist ein funktionsgerechtes Basis- und Befunddokumentationssystem (HAAS et al., 1977) Voraussetzung für eine gute virologische Diagnostik. Dabei fallen viele Daten an, die wegen des erforderlichen Vergleichs lange für einen direkten Zugriff zur Verfügung stehen müssen. Eine manuelle Handhabung der Daten ist umständlich und aufwendig. Deshalb erscheint eine EDV-Lösung sinnvoll (FELDMEIER und REPGES 1976, KAYSER et al., 1976).

Nun wurde jedoch vor kurzem eine EDV-Lösung für die Virologie vorgestellt (BLOMBERG et al., 1979). Sie unterstützt die Laboradministration und Abrechnung. Allerdings ist dafür ein Großrechner mit umfangreichen Ressourcen notwendig. Wir setzten uns dagegen die Aufgabe, ein Patientendaten-Aufnah-

me- und Dokumentationssystem in der medizinischen Virologie unter Zuhilfenahme eines Kleinrechners zu realisieren. Das System hat sich seit zwei Jahren im täglichen Routineeinsatz bewährt (THEUER et al., 1979).

2 Gegebenheiten

Das Institut für Virologie arbeitete für dieses Projekt mit dem Institut für Medizinische Dokumentation, Statistik und Datenverarbeitung der Universität Heidelberg (IDSD) zusammen. Im IDSD stand Hardware, wie in Abbildung 1 dargestellt, zur Verfügung.

Im Routinebetrieb soll das Virologiesystem nur eine Partition (16 k Bytes) des vorhandenen Hauptspeicherplatzes sowie ungefähr 10% der Plattenkapazität nutzen.

Als Systemsoftware steht TSOS zur Verfügung. Das ist ein dialogorientiertes Time-sharing-Betriebssystem.

3 Zielsetzung

Die vor der Arbeit durchgeführte Systemanalyse (BLÜMMEL et al., 1978) ergab, daß eine EDV-Lösung für die Virologie folgende Ziele erfüllen muß:

- Implementation auf Minirechner/Mikrorechner
- Komfortable Datenerfassung und -präsentation im Dialog mit Benutzerführung
- Benutzerfreundliches Suchverfahren mit Modifikationsmöglichkeit aller Daten
- Problemlose Dateiverwaltung
- Direktzugriff über mehrere Aufnahmejahrgänge
- Datenschutz und -sicherung
- Berücksichtigung einer Schnittstelle zu einem Großrechner
- Anfertigung von Statistiken
- Erweiterbarkeit durch Modultechnik

Weiterhin ergibt sich die Aufgabe, das Antwortzeitverhalten bei On-line-Verarbeitung im Time-sharing-Betrieb günstig zu halten.

Um andere Benutzer des Rechners nicht einzuschränken, läuft das System unter Gebrauch von Segmentierungs-Techniken in einer normalen 16 kBytes Partition.

Ein weiteres Problem ergibt sich aus der räumlichen Entfernung zwischen dem Institut für Medizinische Virologie und dem IDSD, dem Standort des Rechners.

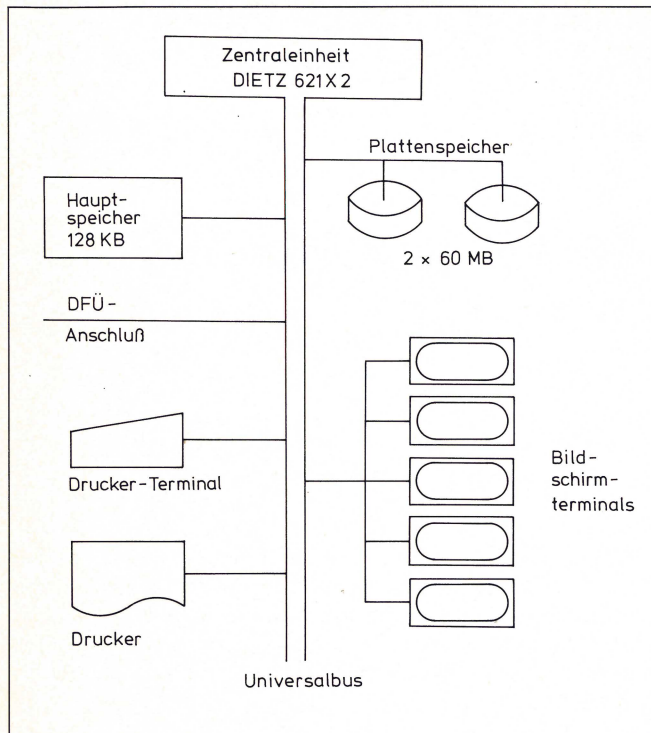
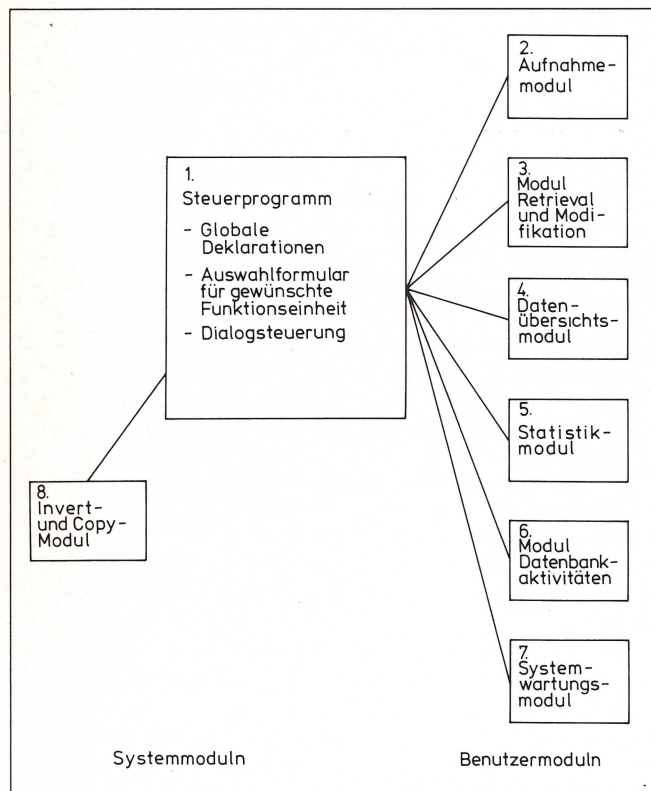


Abb. 1. Rechnerkonfiguration

Schließlich ist die Anwendung des Virologiesystems mit den Systemen der anderen Benutzer des Rechners zu koordinieren. Das Virologie-Programmpaket ist auf einem Mikrorechner implementierbar.

Abb. 2. Programmstruktur



4 Projektdurchführung

4.1 Projektrealisation

4.1.1 Software-Realisation

4.1.1.1 Programmaufbau

Das System gliedert sich in funktionale Moduln. Die Moduln werden von einem Steuerprogramm aufgerufen. Abbildung 2 gibt die realisierte Struktur wieder.

1. Steuerprogramm

Das Steuerprogramm bietet dem Benutzer ein Bildschirmformular dar, anhand dessen er das gewünschte Programm-Modul auswählt (Menütechnik). Das Steuerprogramm sorgt auch durch die Zuordnung Benutzercode – Terminal – Programm für den nötigen Datenschutz.

Wegen der begrenzten Hauptspeicherkapazität ist eine Segmentierung des Programmsystems notwendig. Das Steuerprogramm befindet sich ständig im Hauptspeicher, während jeweils nur ein weiteres Modul in Überlagerungstechnik (Overlay-Technik) den verbleibenden Speicherplatz belegt.

2. Aufnahmemodul

Das Aufnahmemodul dient zur Eingabe der Basisdaten und evtl. bereits vorhandener Diagnosen und Untersuchungsergebnisse unmittelbar nach Eintreffen der Proben und zur Vergabe einer Patientenkenntung. Es bietet Korrekturmöglichkeiten aller einzugebenden Daten; bei Eingabe von verschlüsselten Parametern erfolgt auf Anfrage eine Auflistung der zulässigen Codes.

Da eingegabene Daten nach kurzer Zeit wieder unter mehreren Suchbegriffen zugreifbar sein müssen, ist eine sofortige Invertierung notwendig. Diese bedingt jedoch ein schlechtes Antwortzeitverhalten. Eine temporäre Zwischendatei nimmt daher zunächst alle Patientensätze auf. Nach Abschluß der jeweiligen Sitzung erfolgt automatisch vom System die Übernahme der Daten dieser Datei in die Hauptdatei mit gleichzeitiger Invertierung.

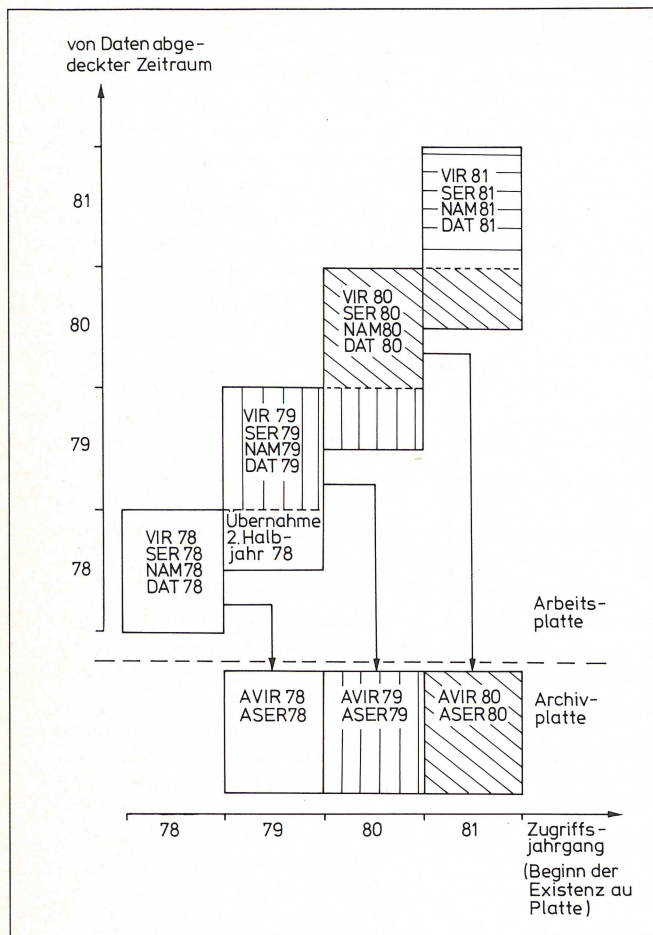
3. Modul Retrieval und Modifikation

Aufgrund der Invertierungen ist ein Patient sofort nach Patientenkenntung oder Name oder Aufnahme datum abrufbar. Auch findet das System Patientengruppen anhand eines Aufnahmezeitraumes oder anhand von Namensmasken. Diese Vorgänge laufen in dem akzeptablen Bereich von weniger als 3 Sekunden ab.

Modifikationen sind die Eingabe neu gewonnener Untersuchungsergebnisse bzw. das Löschen oder Ändern sämtlicher bereits vorhandener Daten oder Teile eines Patientensatzes. Dabei ist u. U. eine Aktualisierung aller invertierten Schlüssel-dateien notwendig.

4. Datenübersichtsmodul

Dieses Modul erstellt einfache Übersichten, wie z. B. eine alphabetische Liste aller Patienten, die einem bestimmten Aufnahmezeitraum angehören und/oder in einem Datenfeld eine bestimmte Ausprägung aufweisen. Außer dem Aufnahmezeitraum, dem Datenfeld und der Ausprägung des Datenfeldes kann der Benutzer noch entscheiden, ob er eine ausführliche Liste mit Titer-Werten der Patienten will oder lediglich eine Liste ohne serologische Untersuchungsergebnisse.



4.1.1.3 Wahl der Programmiersprache

C-BASIC war im Zeitraum der Programmrealisation die einzige am DIETZ-Rechner verfügbare Sprache, die Overlay-Technik, eine komfortable Dateiverwaltung, umfangreiche Textstringoperationen, arithmetische Funktionen und bildschirmorientierte Sonderfunktionen bietet. Sie ermöglicht interaktives Programmieren, da es sich um eine Interpretersprache handelt.

4.2 Systemprofil

Abbildung 6 gibt die Verteilung der Systembenutzerfunktionen auf die verschiedenen Institutionen wieder.

4.3 Projektorganisation

Die drei erstgenannten Autoren bildeten gleichberechtigt das Projektteam. Zunächst erstellten sie in einer gemeinsamen Sitzung einen Zeitplan zur Durchführung des Projekts. Eine Woche war nötig, um gemeinsam die Aufteilung der Aufgabe in Moduln vorzunehmen und ihre Schnittstellen zum Steuerprogramm sowie das Steuerprogramm selbst festzulegen. In einer abschließenden Planungssitzung verteilten wir die Realisation der einzelnen Moduln an die Projektmitglieder unter Berücksichtigung des ausgearbeiteten Zeitplanes. Das zuerst realisierte Steuerprogramm legte alle Aufrufe der Moduln fest. Dies ermöglicht eine koordinierte Erstellung der Moduln. Das Aus-testen und Integrieren der Moduln erfolgt nach dem Top-down-Integrationsverfahren (SCHNUPP, FLOYD 1976, S. 195 ff.).

Einige einfache Spielregeln hielten die Kommunikation und Koordination aufrecht, obwohl die Teammitglieder bei der Arbeitszeiteinteilung frei waren:

Abb. 5 (oben). Dateiwartung und Archivierung

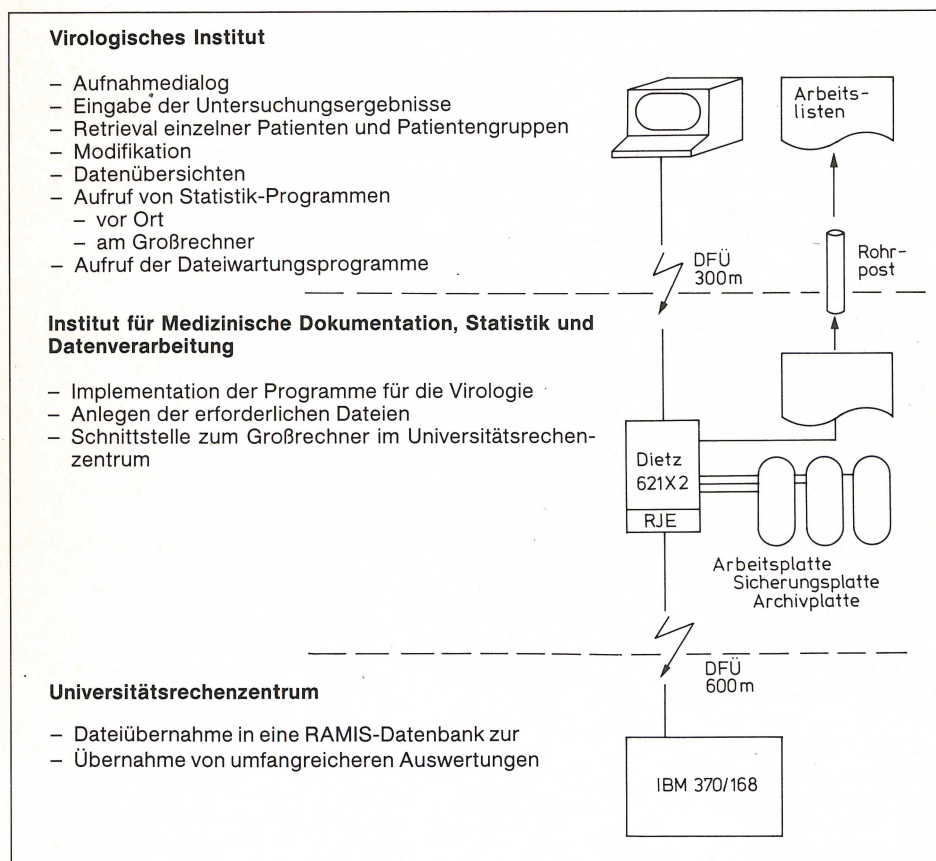


Abb. 6. Systemprofil

Nach jeder interaktiven Programmiersitzung bringt man zwei Dateien auf den neuesten Stand: Die Mitteilungsdatei »NEWS« mit Freitext und eine Programm- und Dateienübersicht in tabellarischer Form. Vor Arbeitsbeginn werden diese Dateien gelesen. Für spätere Zwecke unwichtige Hilfsdateien werden nach Vornamen aus dem Bekanntenkreis benannt. Eine Datei namens »MAX« kann also jederzeit von jedem gelöscht werden. Die Kompilate von Programmen (lauffähige Version der Programme) erhalten den Namen der Quellprogramme, ergänzt durch ein vorangestelltes »C«.

Gleichzeitig zur Programmierung erstellt man die projektbegleitende Dokumentation, und zwar grundsätzlich im Präsens, um später die Gesamtdokumentation ohne weiteres zusammensetzen zu können.

Insgesamt benötigte die Projektdurchführung mit Systemanalyse 8-Mann-Monate Arbeitszeit.

5 Diskussion

Unser virologisches System löst das Problem einer Befunddokumentation für die Virologie. Das Hauptziel ist dabei die Unterstützung der virologischen Diagnostik. Darüber hinaus unterstützt es die virologische Epidemiologie mit Routinestatistiken. Soweit aktuelle Forschungsfragen zu bearbeiten sind, hat sich der Anschluß über Datenfernübertragung an das Universitätsrechenzentrum mit der vorhandenen Datenbank und der Statistik-Software bewährt. Mit diesen Möglichkeiten unterscheidet sich unser System von dem System VIRUS (BLOMBERG et al., 1979). VIRUS unterstützt die Administration eines Labors, das genauso gut auf dem Gebiet der Pathologie tätig sein könnte. Speziell virologische Bedürfnisse befriedigt dieses System offenbar nicht. Darüber hinaus ist das System VIRUS ein Großrechnersystem und setzt damit die Verfügbarkeit eines Großrechners voraus, der nicht im wesentlichen ein Forschungs- und Lehrzwecken dienender Rechner mit den entsprechenden häufigen Systemzusammenbrüchen sein darf. Unser System kann dagegen als isoliertes System laufen, wenn man auf die unmittelbare Möglichkeit der statistischen Auswertung zu Forschungszwecken verzichtet. Die minimale Rechnerkonfiguration, auf der das System implementierbar ist, kostet weniger als 50 000 DM. Die epidemiologischen Routinestatistiken, wie sie auch die staatliche Administration benötigt, bleiben vollständig erhalten. Der Break-even-point dürfte demnach bei der Einsparung einer halben Laborarbeitsstelle liegen.

Unser System befriedigt also virologische Bedürfnisse in moderner Weise. Die vorliegende Arbeit dient jedoch hauptsächlich der Demonstration des Nutzens moderner Software-Entwicklungstechniken.

Die angewandte Top-down-Technik führt zu strukturierter Programmierung und zur Gliederung in sauber definierte Moduln. Hierdurch ist eine einfache Wartung, Änderung und Testung der entwickelten Software garantiert. Für die Gliederung in Moduln haben wir uns an die logische Struktur des Arbeitsablaufs beim Nutzer und die dabei nahegelegte Menütechnik für die Auswahl der Rechnerfunktionen gehalten. Gleichzeitig entspricht die Gliederung in Moduln der an Kleinrechnern erforderlichen Overlay-Struktur.

Nun erscheint jedoch die Top-down-Technik in Lehrbüchern (z. B. SURBÖCK 1978, S. 131 ff.) in Zusammenhang mit der sog. Chiefprogrammer-Organisation. Diese Organisationsform kommt dem Arbeiten in Universitätsinstituten freilich nicht sehr entgegen. Forschung und Entwicklung an Hoch-

schulen muß vielmehr der durch die Lehr- und Ausbildungsfunktion der Universitäten bedingten Fluktuation und dem Individualismus junger, begeisterungsfähiger Mitarbeiter Rechnung tragen. Wie sich in unserer Arbeit zeigte, ist die Top-down-Technik keineswegs nur bei Chiefprogrammer-Organisation anwendbar. Vielmehr erwies sie sich als ideales Hilfsmittel zur Organisation, Koordination und Überwachung einer Teamarbeit. Sie erscheint uns geradezu als Voraussetzung für eine echte, gleichberechtigte Zusammenarbeit mehrerer Individualisten.

Die projektbegleitende Dokumentation durch Anlegen einer Datei mit »NEWS« während der Programmerstellung dient, vom organisatorischen Standpunkt aus betrachtet, der Fortschrittskontrolle, im Interesse der Änderungs- und Wartungsfreundlichkeit eines Software-Produkts der Programmdokumentation und schließlich der Koordination der Teammitglieder. Der letzte Punkt ist besonders wichtig, da nur hiervon eine ausreichende Motivation zum Eintrag und zum Lesen ausgeht. Durch Anordnungen ist hier nichts zu erreichen. In unserem Team wurde der Eintrag so sehr zur selbstverständlichen Gewohnheit, daß ein Teammitglied eines Tages (richtiger: eines Nachts) die Nachricht für seine Kollegen hinterließ, er werde den einzigen Schlüssel zum Rechnerraum mit nach Hause nehmen (zu jener Zeit befanden sich alle Terminals noch im Rechnerraum)!

Das System ist jetzt zwei Jahre in Betrieb. Es hat sich so sehr bewährt, daß es inzwischen auch die immunologischen Schwangerschaftsüberwachungen außerhalb des Virologischen Institutes verwaltet.

Literatur

- RAMIS, User Manual, Mathematical Products Group, Rel. 2.0, Princeton 1979
- BLOMBERG, J., J. MATTSON, B. BÖRJESSIN, S. HAST, P. ANDERSSON, K. HOLM: VIRUS – A Laboratory Information System for Clinical Virology. *Meth. Inform. Med.* **18** (1979) 207–214
- BLÜMMEL, B., U. ECKERT, M. ROTHMUND: CVIRO – Programmpaket für das Institut für Virologie der Universität Heidelberg – Version o.1 vom 1.7.1978
- FELDMEIER, H., R. REPGES: Elektronische Datenverarbeitung in der Medizinischen Mikrobiologie – Leitlinien und kritische Gedanken. *Zbl. Bakt. Hyg., I. Abt. Orig. A* **236** (1976) 422–438
- HAAS, R., H. W. DOERR, E. E. PETERSON, H. SCHMITZ: Über die Vergleichbarkeit virusserologischer Befunde – Ein Beitrag zur Frage des Standardprinzips in der Virusdiagnostik. *Bundesgesundhbl.* **20** (1977) 289–294
- KAYSER, F. H., J. SCHILDKNECHT, J. MUNZINGER, J. P. KARASEK: Ein System der elektronischen Datenverarbeitung für die klinische Mikrobiologie. *Zbl. Bakt. Hyg., I. Abt. Orig. A* **234** (1976) 554–565
- SCHNUPP, P., Ch. FLOYD: Programmentwicklung und Projektorganisation (de Gruyter, Berlin 1976)
- SURBÖCK, K.: Management von EDV-Projekten (de Gruyter, Berlin 1978)
- THEUER, D., H.-U. BURKHARDT, H. W. DOERR: Einführung eines einfachen Systems zur EDV-Befunddokumentation in die Virusdiagnostik. *Bundesgesundhbl.* **22** (1979) 205–212

Eingegangen am 16. 7. 80

Anschriften der Autoren: Priv.-Doz. Dr. med. habil. G. K. Wolf, Institut für Med. Dokumentation, Statistik und Datenverarbeitung der Universität Heidelberg, Im Neuenheimer Feld 305, D-6900 Heidelberg.

EKIP – Ein System zur Verarbeitung registerbezogener medizinischer Daten, insbesondere klinischer Krebsregister*)

A. Altendorf, H. P. Sinn, P. Hermanek, F. P. Gall

Zusammenfassung

EKIP ermöglicht den Aufbau computergestützter klinischer Krebsregister und großer prospektiver Studien. Es handelt sich um ein Programmsystem zur Prüfung und Auswertung von Daten, die in einer speziell dafür entwickelten Datenstruktur gespeichert sind. Leichte Bedienbarkeit, exakte automatische Datenprüfung sowie übersichtliche und leicht lesbare Ausdrücke ermöglichen dem Kliniker maßgebliche Mitarbeit bei der Formulierung der Fragestellungen und bei der Datenauswertung.

Summary

EKIP is a software package for hospital-based cancer registries. It includes different routines for debugging the data and performance of simple statistical calculations. Important for the clinician are: easy mode of access to the data via terminal display, accurate way of checking-up data, self-explanatory print-out. Interested clinicians can participate in working out a model registry that fits his requirements and do simple evaluations by himself.

Einleitung

Angesichts der immer differenter, weil individualisierter werdenden Krebstherapie ist eine vollständige, standardisierte Dokumentation der wichtigsten, die Behandlung beeinflussenden Parameter innerhalb klinischer Krebsregister unbedingt angezeigt. Nur sie ermöglicht in Verbindung mit einer lückenlosen Verlaufsbeobachtung eine fundierte Erfolgskontrolle. Behandlungsmethoden und Therapieergebnisse allein aufgrund persönlicher Erfahrungen zu beurteilen, ist ungenügend. Es ist vielmehr eine systematische Datensammlung und -auswertung erforderlich [1].

Die schier unbegrenzte Speicherkapazität eines Großcomputers ist wie dafür geschaffen, die täglich anfallenden Daten zu speichern. Das Problem liegt in der adäquaten Prüfung und Auswertung dieser Daten. Vorhandene Programmsysteme, wie SPSS [15] wurden für sehr allgemeine statistische Fragestellungen konzipiert. Daher sind sie nur mit relativ großen Vorkenntnissen aufzurufen und liefern für den in der Datenverarbeitung unerfahrenen Anwender schwer verständliche Ausdrücke. So vergrößern allgemeine Bibliotheksprogramme für den Kliniker

die Schwelle zum Computer. Informatiker, Statistiker und andere müssen Mittler sein. Es entstehen leicht Kommunikationsprobleme, die nicht nur zu Verzögerungen, sondern auch zu Mißverständnissen und Fehlinterpretationen der Ergebnisse führen.

In der Abteilung für Klinische Pathologie in der Chirurgischen Universitätsklinik Erlangen wurde in den Jahren 1978 und 1979 ein System entwickelt (Erlanger Krebsregister interaktives Programmsystem), mit dem der Kliniker, vereinfacht gesprochen, seine Fragestellungen direkt in den Computer eintippen kann. Andere Arbeiten, wie Dateierhaltung, Datensicherung und Programmpflege, bleiben weiterhin Fachleuten überlassen. Die Ergebnislisten werden automatisch mit umfangreichen Klartextangaben ausgedruckt und sind somit direkt für den Kliniker lesbar. Dadurch sind Fehlinterpretationen praktisch ausgeschlossen und die Ausdrücke einwandfrei dokumentierbar (selbstdokumentierend).

Seit Mitte 1979 werden alle im Routinebetrieb anfallenden statistischen Fragestellungen mit EKIP bearbeitet. Bisher wurden kolorektale Karzinome [4, 6, 7], kolorektale Polypen [8], Magenfrühkarzinome [14], maligne Melanome [5, 9, 16, 18], Nierenkarzinome [10, 11] und Blasenkarzinome [11] näher mit EKIP analysiert. Die Arbeiten stehen unter Leitung einer Ärztin, die zugleich Diplommathematikerin ist. An Hilfskräften sind derzeit eine angelernte Dokumentationskraft und eine studentische Hilfskraft (zum Ablocken der Bögen) beschäftigt.

Erhebung von Daten

Die Erhebung der Daten für EKIP erfolgt wie allgemein üblich mit standardisierten Datenerhebungsbögen. Die interessierenden Variablen werden der Reihe nach aufgeführt und deren mögliche Ausprägungen numerisch verschlüsselt. Die Art und Gliederung der erhobenen Daten ist für die Verarbeitung mit EKIP unwesentlich. In diesem Zusammenhang kann nicht näher auf die Erstellung standardisierter Erhebungsbögen eingegangen werden, wir verweisen hierbei auf einschlägige Literatur [12, 13]. Es muß jedoch betont werden, daß statistische Erhebungen im allgemeinen und insbesondere in klinischen Krebsregistern außerordentlich sorgfältig geplant werden müssen und die Erhebung ebenso sorgfältig und detailgenau durchzuführen ist. Das Format, in dem die Daten erhoben werden, bleibt über Jahre hinweg konstant, wenn es einmal festgelegt ist. Bereits im Computer gespeicherte Daten nachträglich zu ändern, neue Variablen einzufügen oder die Schlüsselziffern zu verändern, ist mit einem großen Arbeitsaufwand verbunden. Wenn die Erhebung nicht mit der allergrößten Sorgfalt ausgeführt wird, resultiert ein Datenbestand, der mit vielen Fehlern

*) Aus der Abteilung für Klinische Pathologie (Vorstand Prof. Dr. med. P. Hermanek) in der Chirurgischen Klinik (Direktor Prof. Dr. med. F. P. Gall) der Universität Erlangen-Nürnberg

behaftet ist, die nachträglich wieder entfernt werden müssen, wenn sie überhaupt entdeckt werden. Unvollständig erhobene Daten führen in der Auswertung zu wenig relevanten Aussagen und entwerten damit auch die Daten, die vollständig erfaßt sind. Ist z. B. nur von 50 % der erfaßten Patienten bekannt, ob und wann sie ein Tumorrezidiv erlitten haben, so kann aus diesen Informationen keine Aussage über die Rezidivhäufigkeit des Gesamtkrankengutes gewonnen werden.

Gliederung der Daten im Klinischen Krebsregister der Chirurgischen Universitätsklinik Erlangen

Die Daten des Klinischen Krebsregisters werden entsprechend den Gegebenheiten und Erfordernissen einer Chirurgischen und Urologischen Klinik in verschiedenen Organregistern erfaßt (Tab. 1). Dem klinischen Krebsregister ist ein gemeinsam mit der Medizinischen Universitätsklinik geführtes Register kolorektaler Polypen angeschlossen, das formal gleich gestaltet ist.

Der Datensatz jedes Patienten in einem Organregister besteht aus 4 Erhebungseinheiten:

- Stammdaten
- standardisierter WHO-Bogen[19]
- organbezogene klinische und pathologische Daten[12, 13]
- Nachuntersuchungsbefunde

Datenstruktur

Die Datenstruktur für EKIP ist sehr flexibel und gibt dem Benutzer viele Freiheiten bei der Datenerfassung. Sie ermöglicht unter anderem

- die Aktualisierung des bereits vorhandenen Datenbestandes durch neu hinzukommende Patienten und Nachuntersuchungsbefunde bereits gespeicherter Patienten durch einfache Sortiervorgänge;
- die Speicherung beliebig vieler Parameter im Datenbestand eines Patienten; bei einem Auswertungslauf können gleichzeitig maximal 200 Variablen berücksichtigt werden;
- das Anfügen von Spezialerhebungen zu allen Patienten oder zu einer ausgewählten Gruppe eines Registers und die gemeinsame Auswertung mit den Daten dieses Registers;
- die Aufnahme beliebig vieler Nachuntersuchungen in das Register (bei Auswertungen wird automatisch die jeweils letzte Nachuntersuchung eines Patienten berücksichtigt);
- die Auslassung nicht zutreffender Erhebungseinheiten; beispielsweise wird nicht jeder Tumor reseziert, die fehlenden Angaben über das Resektionspräparat beeinflussen den Programmablauf nicht.

Es resultieren beträchtlich unterschiedliche Datensätze mit erfaßten und nicht erfaßten (weil unzutreffenden) Daten, mehrfach erhobenen Nachuntersuchungen, zusätzlichen Spezialuntersuchungen und ebenfalls im Register gespeicherten Stammdaten. Diese Daten sind innerhalb des Datenbestandes eines Patienten nicht nach einer festgelegten Reihenfolge abgespeichert, sondern nach dem Datum ihrer Erhebung. Damit ist für jeden Patienten der Verlauf gut dokumentiert, ähnlich wie in einer Krankengeschichte, in der die Befunde ebenfalls chronologisch abgelegt sind und unzutreffende Einheiten fehlen. Diese Datenstruktur wurde für spätere Verlaufsbeobachtungen konzipiert. EKIP nimmt jedoch zumindest derzeit keine Rücksicht auf die chronologische Reihenfolge der Daten.

Tabelle 1. Gliederung der Daten im Klinischen Krebsregister der Chirurgischen Universitätsklinik Erlangen

kolorektale Malignome
Magenmalignome
maligne Tumoren des Oesophagus
maligne Tumoren von Pankreas und periampullärer Region
maligne Lungentumoren
maligne Melanome
maligne Tumoren der peripheren Weichteile
maligne Schilddrüsentumoren
maligne Knochentumoren
maligne Lymphome und Leukämien
maligne Nierenparenchymtumoren
Nierenbeckentumoren
Harnblasentumoren
maligne Tumoren der Prostata
seltene sonstige maligne urologische Tumoren
andere seltene maligne Tumoren

Tabelle 2. Übersicht über die EKIP – Programme

Daten- prüfung	EKIP-01: Vollzähligkeit der Patienten: Fehlermeldung beim Fehlen definierter Registernummern (die Registernummern werden fortlaufend und lückenlos vergeben)
	EKIP-02: Vollzähligkeit der Erhebungseinheiten: Fehlermeldungen beim Fehlen obligatorischer Erhebungseinheiten und bei Doppelseingabe
	EKIP-03: formale Datenprüfung: Fehlermeldung beim Vorkommen nicht definierter Schlüssel zu den Variablen
	EKIP-04: logische Datenprüfung (Plausibilitätsprüfung): Fehlermeldung beim Vorkommen von nicht erlaubten bzw. logisch unmöglichen Datenkombinationen
	EKIP-05: Vollständigkeit der Nachuntersuchungsbefunde: Fehlermeldung beim Fehlen von Nachuntersuchungsbefunden zu einem definierten Zeitpunkt
Auswertung	EKIP-06: Auszählung der Verteilung aller Variablen; bei metrischen Variablen Errechnung des Mittelwertes
	EKIP-07: Auszählung der Verteilung einer Variablen; bei metrischen Variablen zusätzlich Errechnung von Mittelwert, Standardabweichung und Medianwert
	EKIP-08: Ausgabe der Kenndaten von Patienten (Name, I-Zahl, Archivnummer, Stichtag, Histologienummer) zusammen mit bis zu 10 beliebigen Variablen
	EKIP-12: Ausgabe von Überlebens Tabellen mit alterskorrigierten Überlebensraten
	EKIP-14: Ausgabe von Überlebens Tabellen mit Heilungsraten (NED-Raten, Erscheinungsfreiheitsraten)
	EKIP-15: Ausgabe von Kontingenztafeln zur Klärung von Korrelationen

Anmerkung: Alle Auswertungsprogramme (EKIP-06–15) lassen sich auf beliebig definierte Untergruppen von Patienten eines Registers anwenden.

PROGRAMMBIBLIOTHEK DER ABTEILUNG FUER KLINISCHE PATHOLOGIE IN DER CHIRURGISCHEN UNIV.-KLINIK ERLANGEN
(VORSTAND: PROF. DR. P. HERMANEK)

>>>> START EKIP 6 <<<<

DATUM: 8. 8.80 ZEIT: 14:39 UHR

REGISTER KOLOREKTALER POLYPEN

EKIP 3/80 (SINN, ALTENDORF)
FALL 1200001 BIS 1202556
FALLZAHL: 2201
MONTIERT AM: 8. 8.80
ABSCHLUSSDATUM: 0. 0. 0

Abb. 1. Beispiel für eine
Auszählliste, dargestellt
an den pathohistologischen
Daten im Register kolorek-
taler Polypen (EKIP-06).

32/20	LOKALISATION	30 FLEXURA HEPATICA	59 = 2.7%
		31 TRANSVERSUM	145 = 6.6%
		32 LINKES KOLON	212 = 9.6%
		33 SIGMA	707 = 32.1%
		34 COECUM	49 = 2.2%
		36 ASCENDENS	130 = 5.9%
		37 FLEXURA LIENALIS	64 = 2.9%
		49 REKTUM	821 = 37.3%
		99 UNBEKANNT	13 = 0.6%
32/22	ENTFERNUNG ZUR LINEA DENTATA	998 ENTFAELT, DA POLYP NICHT IM REKTUM	659 = 30.0%
		999 UNBEKANNT	626 = 28.5%
		MITTELWERT (N= 915)	133.42
32/25	HISTOLOGISCHE KLASSIFIKATION	1 ADENOM	1481 = 67.3%
		2 ADENOM MIT SCHWEREN ZELLATYPEN	70 = 3.2%
		3 ADENOM MIT INVASIVEM KARZINOM	157 = 7.1%
		4 POLYPOSES KARZINOM	10 = 0.5%
		5 HYPERPLASTISCHER POLYP	381 = 17.3%
		6 SONST., NICHT-NEOPLASTISCHER POLYP	71 = 3.2%
		7 SONSTIGER BEFUND	19 = 0.9%
		9 UNKLASSIFIZIERT	11 = 0.5%
32/26	ADENOMTYP	1 TUBULAER	1328 = 60.4%
		2 TUBULO-VILLOES	268 = 12.2%
		3 VILLOES	112 = 5.1%
		8 ENTFAELT, DA KEIN ADENOM	485 = 22.0%
		9 UNBEKANNT	7 = 0.3%
32/27	RISIKO BEI ADENOM MIT KARZINOM	1 LOW RISK	120 = 5.5%
		2 HIGH RISK	44 = 2.0%
		8 ENTFAELT, DA KEIN ADENOM MIT CA	2025 = 92.0%
		9 UNBEKANNT	11 = 0.5%
32/28	PSEUDOKARZINOMATOSE INVASION	1 NEIN	2011 = 91.4%
		2 JA	182 = 8.3%
		9 UNBEKANNT	7 = 0.3%
32/29	ENTFERNUNG IM GESUNDEN?	1 SICHER IM GESUNDEN	1269 = 57.7%
		2 NICHT SICHER IM GESUNDEN	143 = 6.5%
		3 NICHT IM GESUNDEN	340 = 15.5%
		9 KEINE AUSSAGE MOEGlich	448 = 20.4%
32/30	MATERIALGEWINNUNG	0 POLYPENSPONTANABGANG	2 = 0.1%
		1 EKTOMIE DURCH SCHLINGE	859 = 39.0%
		2 EKTOMIE DURCH ZANGE	218 = 9.9%
		3 CHIR. EKTOMIE TRANSANAL	50 = 2.3%
		4 REKTOTOMIA POSTERIOR	18 = 0.8%
		5 CHIR. EKTOMIE ABDOMINAL, OHNE WANDEX	26 = 1.2%
		6 CHIR. EKTOMIE ABDOMINAL, OHNE WANDEX	20 = 0.9%
		7 INZISIONSBIOPSIE	352 = 16.0%
		8 SONSTIGE METHODE	590 = 26.8%
		9 UNBEKANNT	65 = 3.0%
32/31	GROESSTER DM DES POLYPENKOPFES	99 UNBEKANNT	63 = 2.9%
		MITTELWERT (N=2137)	10.75
32/33	POLYPENTYP	1 GESTIELT	363 = 16.5%
		2 TAILLIERT	165 = 7.5%
		3 SESSIL	1632 = 74.2%
		9 UNBEKANNT	40 = 1.8%

>>>> STOP EKIP 6 <<<<

PROGRAMMBIBLIOTHEK DER ABTEILUNG FÜR KLINISCHE PATHOLOGIE IN DER CHIRURGISCHEN UNIV.-KLINIK ERLANGEN
(VORSTAND: PROF. DR. P. HERMANEK)

Abb. 2. Beispiel für eine Überlebensstabelle, dargestellt für Patienten im Register der Harnblasentumoren (EKIP-12).

DATUM: 11- 8-80 ZEIT: 8:50 UHR

>>>> START EKIP 15 <<<<

EKIP 3/80 (SINN,ALTENDORF)
FALL 1700001 BIS 1701091
FALLZAHL: 586
MONTIERT AM: 16. 7.80
ABSCHLUSSDATUM: 31.12.79

REGISTER MALIGNER MELANOME

1 = MAENNLICH
2 = WEIBLICH
9 = UNBEKANNT

ZEILEN : 21/18 GESCHLECHT

0	=	???
1730	=	LIPPEN
1731	=	AUGENLIDER
1732	=	AEUSSERES OHR
1733	=	ANDERE GESICHTSTEILE
1734	=	KOPFHAUT UND HALS
1735	=	RUMPF
1736	=	ARM UND SCHULTER
1737	=	BEIN UND HUEFTE
1738	=	MEHREERE REGIONEN
1739	=	UNBEKANNT

SPALTEN: 21/40 TOPOGRAPHIE

GESAMTSUMME: 586
CHI-QUADRATE WERT: 85.7
ANZAHL DER FREIHEITSGRADE: 9

[illegible]

>>>> STOP EKIP 15 <<<<

PROGRAMMBIBLIOTHEK DER ABTEILUNG FUER KLINISCHE PATHOLOGIE IN DER CHIRURGISCHEN UNIV.-KLINIK ERLANGEN
(VORSTAND: PROF.DR.P.-HERMANEK)

>>>> START EKIP 12 <<<< DATUM: 11. 8.80 ZEIT: 9:11 UHR

EKIP 3/80 (SINN,ALTENDORF)
FALL 3300001 BIS 3301086

FALLZAHL: 464
MONTIERT AM: 11. 8.80

ABSCHLUSSDATUM: 31.12.77

REGISTER DER BALKENKARZINOME

U E B E R L E B E N S R A T E N (LFD. NR.: 1)

MON	LAENGE	TOTAL	TOT	AUSG	LEB	EFF	Q	P	F(K)	2 STD-ABW	KORR	F-KORR	2S-KORR
0.0	3.00	I 464-	34-	0-	2-	I 463-0	0.0734	0.9266	I 1.0000	0.0000	0.0000	0.0000	0.0000
3.0	3.00	I 428-	33-	0-	9-	I 423-5	0.0779	0.9221	I 0.9266	0.0242	0.0000	0.0000	0.0000
6.0	3.00	I 386-	24-	0-	13-	I 379-5	0.0632	0.9368	I 0.8544	0.0329	0.0000	0.0000	0.0000
9.0	3.00	I 349-	19-	0-	12-	I 343-0	0.0554	0.9446	I 0.8003	0.0375	0.0000	0.0000	0.0000
12.0	6.00	I 318-	20-	0-	40-	I 298-0	0.0671	0.9329	I 0.7560	0.0406	0.9591	0.7883	0.0423
18.0	6.00	I 258-	12-	0-	22-	I 247-0	0.0486	0.9514	I 0.7053	0.0437	0.0000	0.0000	0.0000
24.0	6.00	I 224-	11-	0-	20-	I 214-0	0.0514	0.9486	I 0.6710	0.0459	0.9175	0.7314	0.0500
30.0	6.00	I 193-	10-	0-	11-	I 187-5	0.0533	0.9467	I 0.6365	0.0480	0.0000	0.0000	0.0000
36.0	6.00	I 172-	6-	0-	18-	I 163-0	0.0368	0.9632	I 0.6026	0.0500	0.8754	0.6883	0.0571
42.0	6.00	I 148-	6-	0-	9-	I 143-5	0.0418	0.9582	I 0.5804	0.0513	0.0000	0.0000	0.0000
48.0	12.00	I 133-	6-	0-	25-	I 120-5	0.0498	0.9502	I 0.5561	0.0529	0.8330	0.6676	0.0635
60.0	12.00	I 102-	7-	0-	29-	I 87-5	0.0800	0.9200	I 0.5284	0.0549	0.7905	0.6685	0.0694
72.0	12.00	I 66-	5-	0-	21-	I 55-5	0.0901	0.9099	I 0.4862	0.0591	0.0000	0.0000	0.0000
84.0	12.00	I 40-	2-	0-	11-	I 34-5	0.0580	0.9420	I 0.4424	0.0654	0.0000	0.0000	0.0000
96.0	12.00	I 27-	2-	0-	11-	I 21-5	0.0930	0.9070	I 0.4167	0.0710	0.0000	0.0000	0.0000
108.0	12.00	I 14-	0-	0-	4-	I 12-0	0.0000	1.0000	I 0.3779	0.0829	0.0000	0.0000	0.0000
120.0	12.00	I 10-	0-	0-	4-	I 8-0	0.0000	1.0000	I 0.3779	0.0829	0.5831	0.6482	0.1422
132.0	*****	I 6-	1-	0-	5-	I 3-5	0.2857	0.7143	I 0.3779	0.2686	0.0000	0.0000	0.0000

>>>> STOP EKIP 12 <<<<

Datenprüfung

Voraussetzung eines exakten und aussagekräftigen Datenbestandes ist eine Prüfung der Daten. Sie läßt sich in drei logisch aufeinanderfolgende Schritte unterteilen:

- 1) Prüfung auf Vollzähligkeit
- 2) Prüfung auf formale Fehler (nicht definierte Schlüssel)
- 3) Prüfung auf logische Fehler (Plausibilitätsfehler, nicht zueinander passende Schlüssel)

Hierfür wurden die EKIP Programme 01 bis 05 entwickelt (Tab. 2). Trotz aller Sorgfalt bei der Datenerhebung, auch bei Datenerhebung durch mehrere Personen unabhängig voneinander, treten formale und logische Fehler in erstaunlich hohem Prozentsatz auf. Daher erfordert jede Datenauswertung eine sehr sorgfältige programmierte Datenprüfung. Die in EKIP integrierten Prüfprogramme sind daher wesentliche Voraussetzung einer verlässlichen Auswertung.

Auswertung

In der klinischen Krebsstatistik treten vor allem drei Fragestellungen auf:

- die Frage nach der Häufigkeit von bestimmten Merkmalen und Merkmalsgruppen
- die Frage nach Korrelationen zwischen bestimmten Merkmalen
- die Frage nach der Kurz- und Langzeit-Prognose

Die Beantwortung dieser Fragestellungen erfolgt mittels verschiedener Programme im EKIP-System (Tab. 2), wobei vorerst auf die Darstellung absoluter und relativer Häufigkeiten mehr Wert gelegt wurde als auf die Anwendung statistischer Testverfahren.

Das zentrale Auswertungsprogramm ist EKIP-06. Es erzeugt eine »Auszählliste« der vorgekommenen Schlüssel aller Variablen für die Patienten eines Registers und liefert damit einen guten Überblick über die erfaßten Daten (Abb. 1). Davon ausgehend können mit den anderen Auswertungsprogrammen die oben erwähnten Fragestellungen bearbeitet werden.

Die Errechnung von Überlebensraten (EKIP-12) und Heilungsraten (EKIP-14) erfolgt entsprechend den Empfehlungen der UICC[17] und des American Joint Committee for Cancer Staging and End Results Reporting[2] nach dem von Berkson und Gage[3] angegebenen Verfahren (actuarial method). Die Überlebensraten werden in Form einer life-table ausgegeben. So ist dann auch die graphische Darstellung in Form von Überlebenskurven leicht möglich. Zu den beobachteten 1-, 2-, 3-, 4-, 5- und 10-Jahres-Überlebensraten werden die zugehörigen alterskorrigierten Raten berechnet (Abb. 2). Weitere Auswertungsprogramme umfassen die Bildung von Kontingenztafeln (EKIP-15) (Abb. 3), Auszählung der Ausprägungen einer Variablen (EKIP-07) und Ausgabe der Kenndaten von Patienten (Name, Geburtsdatum, Stichtag, u. a.) zusammen mit anderen erfaßten Schlüsseln dieser Patienten (EKIP-08).

Sämtliche Programme sind im Bildschirmdialog oder, zeitsparender bei umfangreichen Auswertungen, im Stapelbetrieb aufrufbar. Dabei werden höchstens zwei Steuerkarten pro Aufruf benötigt.

Alle Auswertungsprogramme können auch beliebig definierte Teilkollektive eines Registers bearbeiten. Die Definition dieser Gruppen erfolgt anhand von bis zu 20 verschiedenen Be-

dingungen. Mehrere Bedingungen können durch Konjunktion (»UND«) oder Disjunktion (»ODER«) miteinander verknüpft werden.

Programmtechnische Details

EKIP wurde in der Programmiersprache FORTRAN auf der Rechenanlage TR 440 entwickelt und ist dort für das klinische Krebsregister der Chirurgischen Universitätsklinik Erlangen im Einsatz. Das verwendete FORTRAN 440 gleicht weitgehend dem ANSI 3.9-Fortran (FORTRAN 77). Die Übertragung von EKIP auf Rechenanlagen, die einen Compiler mit diesem Standard besitzen, dürfte keine größeren Schwierigkeiten bereiten; es sind jedoch einige spezielle zusätzliche Assembler-Routinen erforderlich.

Für die Verarbeitung registerbezogener Daten mit EKIP gelten folgende Einschränkungen: Die Daten müssen in der vorgegebenen Datenstruktur vorliegen, d. h. die Spalten 1–17 aller Datensätze sind belegt mit Registerart, Patientennummer, Datum der Erhebung und Erhebungsart. Die Spalten 18–80 enthalten die eigentlichen Informationen. Als Variablen sind nur Integer-Zahlen zugelassen, die maximal vierstellig sind. Metrische Variablen müssen auch als Integer-Größen erfaßt und verarbeitet werden. Es können maximal 99 Register gebildet werden. Jedes Register kann höchstens 99999 Fälle mit jeweils 999 unterschiedlich formatierten Sätzen enthalten. Davon können jedoch nur 20 unterschiedlich formatierte Sätze mit insgesamt 200 Variablen gleichzeitig ausgewertet werden. Diese Begrenzungen stellen jedoch in der Praxis keine Einschränkungen dar.

Schlußbemerkungen

EKIP wurde speziell für den Aufbau computergestützter klinischer Krebsregister und für große onkologische prospektive Studien entwickelt. Die Daten sind in einer speziellen Datenstruktur gespeichert. EKIP ist in Hinblick auf eine maßgebliche Mitarbeit des klinischen Anwenders konzipiert. Daher wurde besonderes Gewicht auf 3 Punkte gelegt:

1. leichte Bedienbarkeit,
2. exakte automatische Datenprüfung,
3. übersichtliche und leicht lesbare Ausdrücke.

Der Kliniker kann sich bei der Formulierung der Fragestellungen wie auch bei der Datenauswertung aktiv beteiligen. Die an vielen Orten in der Praxis bestehenden Kommunikationsprobleme zwischen Klinik und Datenverarbeitung werden so beträchtlich reduziert.

EKIP ist an jedem Großcomputer implementierbar und arbeitet sowohl im benutzerfreundlichen Bildschirmdialog als auch im schnelleren Stapelbetrieb. Die Programmaufrufe sind leicht erlernbar, die Ausdrücke sind nicht verschlüsselt, leicht lesbar und damit unmittelbar verständlich. Die für EKIP erforderliche Datenstruktur ist registerbezogen. EKIP ist für klein- wie großkomplizierte strukturierte Register geeignet.

Literatur

- [1] American College of Surgeons: Cancer Registry Manual. Chicago, 1974
- [2] American Joint Committee for Cancer Staging and End-Results Reporting: Manual for Staging of Cancer. American Joint Committee, Chicago, 1977

Abb. 3 (linke Seite). Beispiel für eine Kontingenztafel, dargestellt an der Verteilung der Lokalisation maligner Melanome in Abhängigkeit vom Geschlecht der Patienten (EKIP-15).

- [3] BERKSON J., GAGE R. P.: Calculation of survival rates for cancer Proc. Staff Meet. Mayo Clin. 25: 170-286 (1950)
- [4] GALL F. P., HERMANEK P.: Therapie des Rektumkarzinoms Dtsch. Ärzteblatt (1980) In Druck.
- [5] GALL F. P., TONAK J.: Die chirurgische Therapie des malignen Melanoms. In: TONAK J., WEIDNER F. (Hrsg.): Das maligne Melanom der Haut. perimed-Verlag, Erlangen, 1980
- [6] HERMANEK P.: Pathomorphologische Aspekte zu kontinenzerhaltenden Therapieverfahren beim Mastdarmkrebs. In: Reifferscheid M., Langer S. (Hrsg.): Der Mastdarmkrebs. Thieme, Stuttgart, 1980
- [7] HERMANEK P., GALL F. P., ALTENDORF A.: Prognostic groups in colorectal carcinoma. Oncology. In Druck
- [8] HERMANEK P.: Kolorektale Polypen und Polyposen: klinisch-relevante pathologisch-anatomische Grundlagen. In: Schwaiger M. (Hrsg.): 12. Freiburger Chirurgengespräche 1980. In Druck.
- [9] HERMANEK P.: Malignes Melanom. Histologie und Klassifikation. In: TONAK J., WEIDNER F. (Hrsg.): Das maligne Melanom der Haut. perimed-Verlag, Erlangen, 1980
- [10] HERMANEK P., SIGEL A., CHLEPAS S.: Grading and Staging in Renal Cell Cancer. Pathol. Reseach and Practice 165, 103 (1979)
- [11] HERMANEK P.: Urologische Tumorpathologie – Methoden, Aufgaben, Ergebnisse –. Wien. Klin. Wschr. 1980. In Druck.
- [12] HERMANEK P., GALL F. P.: Grundlagen der klinischen Onkologie, G. Witzstrock, Baden-Baden – Köln – New York, 1979
- [13] HERMANEK P., GALL F. P.: Lungentumoren. G. Witzstrock, Baden-Baden – Köln – New York, 1979
- [14] MÜHE E., HERMANEK P.: Therapie-Ergebnisse beim Magenfrühkarzinom. In: BEGER H. G., BERGEMANN W., OSHIMA H. (Hrsg.): Das Magenkarzinom. Thieme, Stuttgart, 1980
- [15] NIE N. H., STEINBRENNER K., BENT D. H., JENKINS J. G., HULL C. H., BEUTEL P., KÜFFNER H., RÖCK E., SCHUBÖ W.: SPSS 7 Statistik-Programm – System für die Sozialwissenschaften, Gustav Fischer Verlag, Stuttgart – New York, 1980
- [16] TONAK J., HOFERICHTER S., ALTENDORF A.: ERLANGER Therapie-schemata – Grundlagen, Ergebnisse und Behandlung von Rezidiven. In: TONAK J., WEIDNER F. (Hrsg.): Das maligne Melanom der Haut. perimed-Verlag, Erlangen, 1980
- [17] UICC: TNM General Rules. 2nd ed. UICC, Geneva, 1974
- [18] WEIDNER F., ALTENDORF A., NEUMÜLLER G.: Metastasierungs-muster maligner Melanome. In: Tonak J., Weidner F. (Hrsg.): Das maligne Melanom der Haut. perimed-Verlag, Erlangen 1980
- [19] WHO: Handbook for Standardized Cancer Registries. World Health Organisation, Geneva, 1976

Eingegangen am 6. 8. 80

Anschrift der Verfasser: Dr. Annelore Altendorf, Abt. f. Klinische Pathologie, Chirurgische Universitätsklinik, Maximiliansplatz, 8520 Erlangen

STATISTISCHE VERFAHREN

EDV in Medizin und Biologie 11 (4), 108–112, ISSN 0300-8282

© Eugen Ulmer GmbH & Co., Gustav Fischer Verlag KG, Stuttgart

Erfahrungen mit kubischen Spline-Funktionen bei der Konstruktion altersspezifischer Referenzbereiche

W. Rehenning¹⁾, J.-D. Ringe²⁾, J. Berger¹⁾

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der Konstruktion altersspezifischer Referenzbereiche anhand von Daten, die durch radiologische Messungen gewonnen wurden. Die Struktur der Daten war so, daß durch keine Transformation eine Linearisierung des Zusammenhangs zwischen dem Alter und den Meßgrößen zu erreichen war, so daß Methoden der linearen Regressionsrechnung nicht anwendbar waren. Daher wurde die Regression sowohl durch Ausgleichspolynome bis zum achten Grade als auch durch glättende, kubische Spline-Funktionen geschätzt. Außerdem wurden altersabhängige Referenzbereiche zu einer vorgegebenen Sicherheitswahrscheinlichkeit konstruiert.

Die Brauchbarkeit der Referenzbereiche wurde anhand der zur Konstruktion benutzten Meßpunkte und anhand einer unabhängigen Stichprobe von $n = 84$ Probanden heuristisch überprüft.

Es zeigte sich, daß der glättende Spline den Polynomen zur Wiedergabe der Mittelwerte überlegen ist, da er diese gut schätzt, ohne den bei Polynomen höheren Grades auftretenden Wellenverlauf zu zeigen. Zur Konstruktion des Regressionsgürtels um den Spline mußten jedoch Polynome benutzt werden.

Summary

This work is concerned with the construction of age specific reference ranges with the help of data gained by radiological measurements. The structure of the data allowed no transformation to linearize the connection between age and the measured quantities so that methods of linear regression could not be

¹⁾ Abteilung für Mathematik in der Medizin (Leiter: Prof. Dr. J. Berger)

²⁾ Abteilung für Osteologie (Leiter: Prof. Dr. F. Kuhlencordt), Universitätskrankenhaus Hamburg-Eppendorf

applied. Therefore, the regression equation was estimated by polynomials up to the eighth degree as well as by smoothing cubic spline functions. Furthermore, age dependent reference ranges were constructed to a given inclusion probability.

The applicability of the method was checked empirically with the help of the measured data used for the construction and, in addition, by the values of an independent sample of $n = 84$ persons.

It proved that the spline functions were superior to the polynomials in estimating the regression function because they showed no oscillations as the polynomials of higher degree. For the construction of the regression belt about the spline function, however, the use of polynomials was indispensable.

Einleitung

Viele biologische Parameter hängen in komplexer Form vom Alter ab.

Oft gelingt es durch eine geeignete Transformation, eine Linearisierung des Zusammenhangs zwischen Alter und Zielparаметer zu erreichen und danach mit den Mitteln der linearen Regressionstheorie entsprechende Referenzbereiche zu berechnen. Gelingt eine solche Linearisierung nicht, so ist es schon schwieriger, eine geeignete Regressionsfunktion und vor allem entsprechende Regressionsgürtel zu finden, die die folgenden Eigenschaften haben:

1. Die Regressionsfunktion in bezug auf das Alter soll möglichst einfach sein und keine systematischen Abweichungen von der wahren Regression zeigen.
2. Der Referenzbereich soll einen bestimmten Prozentsatz (z. B. 95 %) aller Werte der Referenzpopulation einschließen (1).

Nachstehend werden verschiedene mathematische Ansätze zur Konstruktion derartiger Referenzbereiche heuristisch auf ihre Brauchbarkeit untersucht und die Ergebnisse kritisch miteinander verglichen.

Material und Methoden

Untersuchungsgut und Meßmethode

Das Untersuchungsgut besteht aus 705 Probanden – Schulkinder, Studenten, Klinikpersonal, Patientenangehörigen und teilweise auch Patienten mit leichten, skelettunabhängigen Erkrankungen – im Alter zwischen 6 und 94 Jahren, 379 weiblichen und 326 männlichen, bei denen aufgrund anamnestischer Kriterien eine lokalisierte oder generalisierte Skeletterkrankung bzw. eine andere Erkrankung mit möglicher Auswirkung auf das Skelettsystem ausgeschlossen wurden, die zur Schätzung der Referenzbereiche dienen und 84 Probanden (42 Männer und 42 Frauen), die zur Validierung dieser Methoden herangezogen wurden.

Zur direkten Messung des Knochenmineralgehaltes wurde die ^{125}Jod -Photonenabsorptionstechnik nach CAMERON und SØRENSEN (2) angewendet. Hierbei wird in einem Scanvorgang die vom Knochengewebe absorbierte monochromatische Strahlung des Isotops ^{125}J gemessen. Nach entsprechender Geräteeichung kann der angeschlossene Prozeßrechner aus der gemessenen Absorptionskurve den Knochendurchmesser (cm) und den Mineralgehalt (g/cm) berechnen (CAMERON und Mitarb. (3), KUHLENCORDT und Mitarb. (4), v. ROTH und Mitarb. (5) und RINGE und Mitarb. (6)).

Die Messungen wurden – getrennt für Radius und Ulna – jeweils am linken Unterarm bzw. bei Linkshändern am rechten

Unterarm an zwei Meßorten – 1/10 bzw. 1/3 der Distanz zwischen Processus styloideus ulnae und Olecranon von distal abgetragen – vorgenommen. Am 1/3 – Meßort beider Unterarmknochen wird praktisch ausschließlich kortikales Knochengewebe gemessen, am distalen 1/10 – Meßort wird auch spongioser Knochen erfaßt (SCHLENKER und v. SEGGEN (7)). Als Maß für die Dichte des Knochens wurde noch der Quotient Mineralgehalt/Knochendurchmesser (g/cm²) gebildet.

Statistische Methode

1. Schätzung der Erwartungswerte

Da der Mineralgehalt des Knochens einer Person etwa bis zum 20. Lebensjahr zunimmt, dann über eine gewisse Zeitspanne konstant bleibt, um dann im Alter wieder abzunehmen, ist es nicht möglich, durch eine Transformation der Variablen eine Linearisierung dieses Zusammenhangs zu erreichen. Daher wurden zur mathematischen Beschreibung dieser nichtlinearen Regression die folgenden beiden Modellansätze gewählt:

1. Ausgleichspolynome bis zum achten Grad

Im gesamten Altersbereich wurde die Regression durch ein Polynom bis zum achten Grad geschätzt:

$$y_i = \sum_{j=0}^8 a_j x_i^j + \varepsilon_i$$

Dabei ist y_i der Meßwert der i -ten Person, x_i ist das Alter der Person und ε_i ist ein Fehlerglied ($i = 1, \dots, n$). Die Berechnung der Konstanten a_j wird durch die Minimalbedingung:

$$\text{SAQ} = \sum_{i=1}^n \varepsilon_i^2 = \text{Minimum}$$

ermöglicht.

Hierbei wurde schrittweise nach aufsteigenden Polynomgraden verfahren.

2. Glättende Splinefunktionen

Es wurde ein natürlicher, ausgleichender, kubischer Spline nach der Methode von H. SPÄTH berechnet (8). Bezeichnet man die Altersabszissen der Größe nach geordnet mit x_i ($i = 1, \dots, r$), ist also:

$$x_1 < x_2 < \dots < x_r,$$

so wird in dem Intervall $x_i \leq x \leq x_{i+1}$

$$f_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i$$

gesetzt. Die Berechnung der unbekannten Konstanten a_i, b_i, c_i und d_i läßt sich erreichen, wenn man fordert, daß die Ordinatendifferenzen zwischen der Splinefunktion und den Mittelwerten der Ordinaten der Meßwerte bei festen x_i den Sprüngen der Splinefunktion in der dritten Ableitung positiv proportional sind, wenn also für positive Gewichte p_i gilt:

$$p_i(f_i(x_i) - \bar{y}_i) = t_i,$$

wobei

$$t_i = f_i'''(x_i) - f_{i-1}'''(x_i)$$

und an den Rändern

$$t_1 = f_1'''(x_1) \text{ und } t_r = -f_{r-1}'''(x_r)$$

gesetzt wird. Die hierbei auftretenden Gewichte p_i sind Faktoren, die lokal den Grad der Glattheit des Splines bestimmen. Je kleiner sie gewählt werden, desto glatter wird der Spline und je größer sie gewählt werden, desto mehr bekommt der Spline interpolatorischen Charakter. Durch Umformung erhält man ein fünfdiagonales, symmetrisches Gleichungssystem mit positiv definiter Koeffizientenmatrix (9), das mit Hilfe eines von SPÄTH (8) angegebenen FORTRAN-Programmes gelöst wurde. Das Programm sieht zunächst vor, daß für jeden Abszissenwert genau ein Ordinatenwert vorliegt. Daher wurden die zu einer Abszisse x_i gehörenden Ordinatenwerte durch ihren Mittelwert \bar{y}_i ersetzt und die Gewichte p_i entsprechend der Anzahl der zu x_i gehörenden Ordinaten erhöht. Der Ausgangsgewichtsvektor \bar{p} wurde empirisch so lange verändert, bis die Regressionskurve hinreichend glatt verlief.

II. Bestimmung der Quantilen

Um die Abweichungen der Einzelwerte um die Regressionskurve darstellen zu können, wurde eine »Streufunction« $s(x)$ folgendermaßen bestimmt: Diese Abweichungen zwischen beobachteten und erwarteten Werten wurden quadriert und die Quadrate in Abhängigkeit vom Alter aufgetragen. Durch die so erhaltenen Punkte wurde ebenfalls eine glatte Ausgleichskurve nach den unter I und II angegebenen Methoden gelegt. Die altersabhängige Streufunktion $s(x)$ wird dann als Quadratwurzel aus der so erhaltenen Funktion definiert. Einen empirischen »Referenzbereich« kann man dann folgendermaßen bekommen: Soll der Bereich durch die Einschlußwahrscheinlichkeit $1-\alpha$ (z. B. 0,95) gekennzeichnet sein, so bildet man die beiden Funktionen:

$$g_1(x) = r(x) - t_\gamma s(x)$$

$$\text{und } g_2(x) = r(x) + t_\gamma s(x)$$

Hierbei ist $r(x)$ die geschätzte Regressionsfunktion. Dann sollen alle Punkte (x,y) mit $g_1(x) \leq y \leq g_2(x)$ den gewünschten Normbereich bilden. Hierbei ist t_γ mit $\gamma = 1 - \alpha/2$ das entsprechende Fraktile der Standardnormalverteilung. Da dieses Verfahren im Prinzip die Normalverteilung der Meßwerte um die Regressionsfunktion voraussetzt, muß man unter Umständen vor der Rechnung die Ordinaten in geeigneter Weise transformieren. Im vorliegenden Datenmaterial wurde mit den Logarithmen der Meßwerte gearbeitet.

III. Kriterien für die Beurteilung der Verfahren

Weil eine theoretische Begründung für diese Vorgehensweise nicht ohne weiteres gegeben werden kann, müssen die vorgeschlagenen Verfahren am Datenmaterial empirisch auf ihre Brauchbarkeit überprüft werden. Dazu wurden folgende Kriterien benutzt:

- a) Die Regressionsfunktion soll möglichst glatt sein und darf die Mittelwerte nicht systematisch verfehlen.
- b) Der Referenzbereich soll 95 Prozent der den Gürtel definierenden Punkte enthalten und sollte möglichst schmal sein.
- c) Die Zahl der außerhalb des Referenzbereiches liegenden Werte sollte etwa fünf Prozent betragen und diese Punkte sollten sich gleichmäßig, d. h. unabhängig vom Alter um den Referenzbereich verteilen.
- d) Für neu hinzukommende Punkte, die nicht zur Definition des Referenzbereiches herangezogen wurden, sollen die Forderungen a) bis c) gelten.

Ergebnisse

I. Vergleich der Regressionsmodelle

In der Abbildung 1 sind die Messungen der Knochendichte bei weiblichen Personen am distalen 1/10-Meßort und die mittels der Spline-Funktion bzw. der Polynomregression erwarteten Schätzungen eingezeichnet.

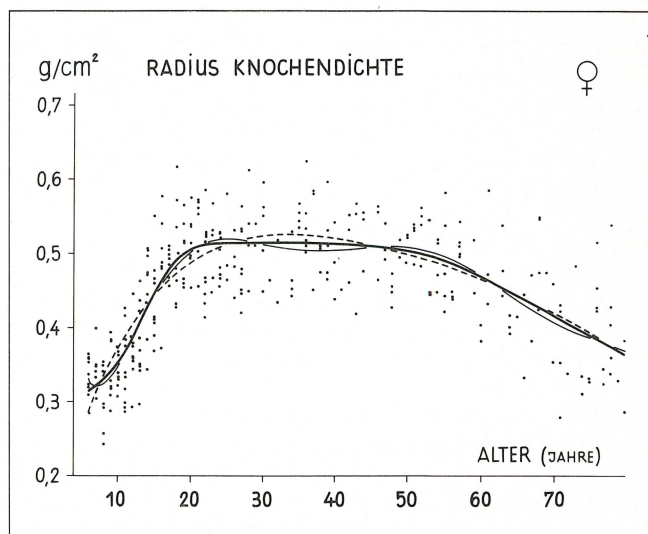
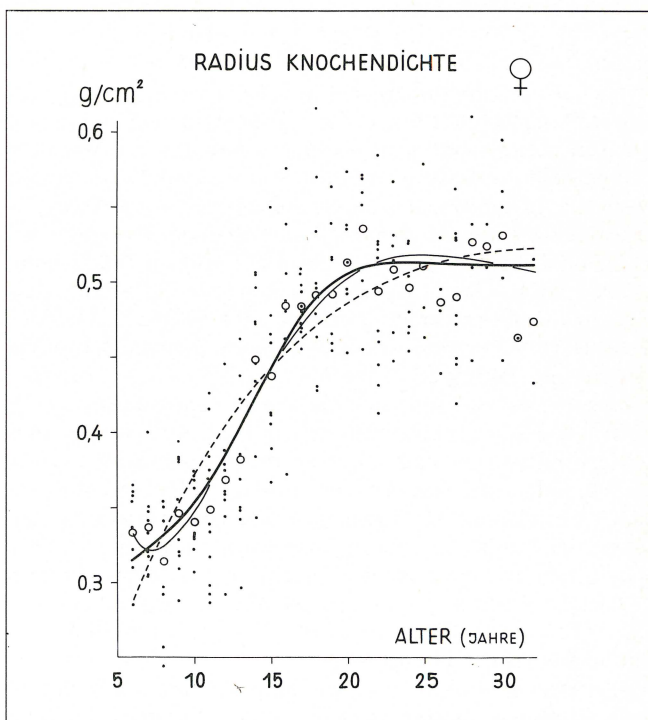


Abb. 1. Gegenüberstellung der mittels Spline-Funktion (dicke Linie), Polynom fünften Grades (gestrichelte Linie) und Polynom achten Grades (dünne Linie) geschätzten und tatsächlich beobachteten Werte der Knochendichte bei weiblichen Personen.

Abb. 2. Gegenüberstellung der beobachteten Werte der Knochendichte, der Mittelwerte pro Altersklasse (Kreise), der Regressionswerte anhand der Spline-Funktion (dicke Linie), des Polynoms fünften Grades (gestrichelte Linie) und des Polynoms achten Grades (dünne Linie). Vergrößerter Ausschnitt aus Abb. 1.



Obwohl alle Kurven den Verlauf der Regression im großen recht gut wiedergeben, zeigt sich doch, daß der steile Anstieg der Knochendichte zwischen etwa dem 10. und 18. Lebensjahr infolge des Wachstumsschubes durch das Polynom fünften Grades nicht befriedigend wiedergegeben wird. Das zeigt sich besonders deutlich in Abbildung 2, in der die Kurvenverläufe aus Abbildung 1 in vergrößertem Maßstab dargestellt sind.

Es ist klar ersichtlich, daß das Polynom fünften Grades die Mittelwerte – Kreise – systematisch verfehlt. Obwohl das Polynom achten Grades die Mittelwerte besser darstellt, ergibt sich doch eine sehr störende Welligkeit über den gesamten Verlauf, die dem Datenkörper nicht angemessen ist. Diese Welligkeit würde bei einer Erhöhung des Polynomgrades noch zunehmen, da man schließlich zu einem stark oszillierenden Interpolationspolynom durch die Mittelwerte kommen würde. Diese Nachteile lassen sich bei der glättenden Spline-Funktion vermeiden; für eine befriedigende Darstellung des Wachstumsschubes mußte jedoch mit altersabhängigen Ausgangsgewichten – unter 24 Jahren mit $p_i = 0,003$, darüber mit $p_i = 0,001$ – gerechnet werden. Ein analoges Verhalten der Spline-Funktion und der Polynomregressionen ergab sich auch bei den übrigen Variablen, auf deren Darstellung hier verzichtet sei.

II. Abgrenzung der Quantilen

Obwohl die Spline-Funktionen bei der Schätzung der Regression brauchbar waren, erwies es sich jedoch als nicht praktikabel, befriedigende Streufunktionen $s(x)$ mit ihrer Hilfe abzuleiten, da kein Ausgangsgewichtsvektor p gefunden werden konnte, der die Streufunktion ohne störende Welligkeit schätzte. Daher wurden die Streufunktionen mit Hilfe von Ausgleichspolynomen zweiten bis dritten Grades berechnet, wobei sehr befriedigende Darstellungen resultierten.

Um Punkt c) der gestellten Kriterien erfüllen zu können, erwies es sich als notwendig, mit den Logarithmen der Meßwerte zu arbeiten, da die Verteilung der Werte um die Regression bei nicht logarithmierten Werten nach oben schief ist. Diese Logarithmierung der Ordinaten führt natürlich im Originalmaßstab zu unsymmetrischen Referenzbereichen, wie aus den Abbildungen 3 und 4 zu sehen ist, in denen der Knochenmineralgehalt bei Männern bzw. Frauen am 1/10 – Meßort mit den zugehörigen Referenzbereichen eingezeichnet sind.

Wie man sieht, liegen von den 326 bei männlichen Probanden gemessenen Werten insgesamt 12 – 5 oberhalb, 7 unterhalb – außerhalb des Referenzbereiches; von den 379 bei weiblichen Personen gemessenen Werten sind es 17 – 7 oberhalb, 10 unterhalb –, wobei dieses »Herausfallen« dem erwarteten Anteil von fünf Prozent entspricht und die herausfallenden Punkte annähernd gleichmäßig über die Altersgruppen verteilt sind. Um das Erfülltsein der in Punkt c) aufgestellten Forderung an die Referenzbereiche besser beurteilen zu können, wurden für vier zufällig herausgegriffene Kurven 50%–Gürtel berechnet, d. h. es wurde mit dem Fraktile $t_{0,5} = 0,6745$ gearbeitet.

Gleichzeitig wurden, entsprechend der Lage auf der Altersachse, fünf etwa gleich große Gruppen gebildet und pro Gruppe ausgezählt, wie viele Meßwerte jeweils unter bzw. über den 50% – Gürtel fielen. Die Ergebnisse sind in der Tabelle 1 dargestellt. Bei Vorliegen von Normalverteilungen müßten die Wahrscheinlichkeiten, um in eine der Zeilen zu fallen, $P = 1/4$, $1/2$ bzw. $3/4$ betragen. Diese Hypothesen liefern einen formalen χ^2 -Test mit 10 Freiheitsgraden. Diese χ^2 -Werte sind ebenfalls in der Tabelle 1 aufgeführt.

Gleichzeitig wurde zur Beurteilung der Breite des Regressionsgürtels jeweils die Summe der Abweichungsquadrate mit

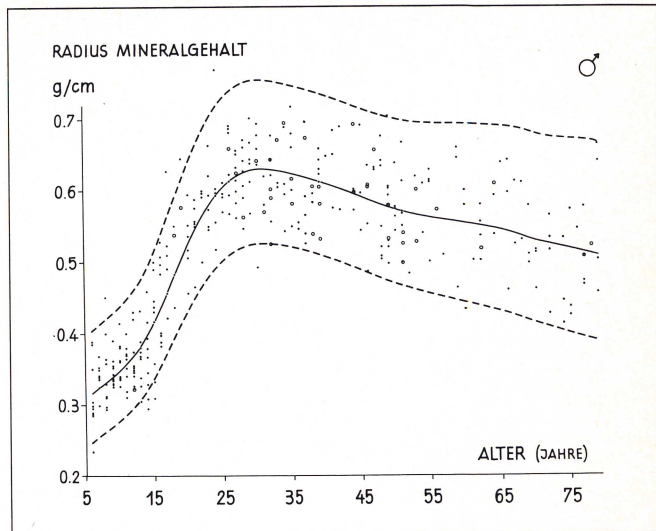


Abb. 3. Beobachteter Mineralgehalt bei männlichen Probanden und 95%-Referenzbereich für den Mineralgehalt

Abb. 4. Beobachteter Mineralgehalt bei weiblichen Probanden und 95%-Referenzbereich für den Mineralgehalt

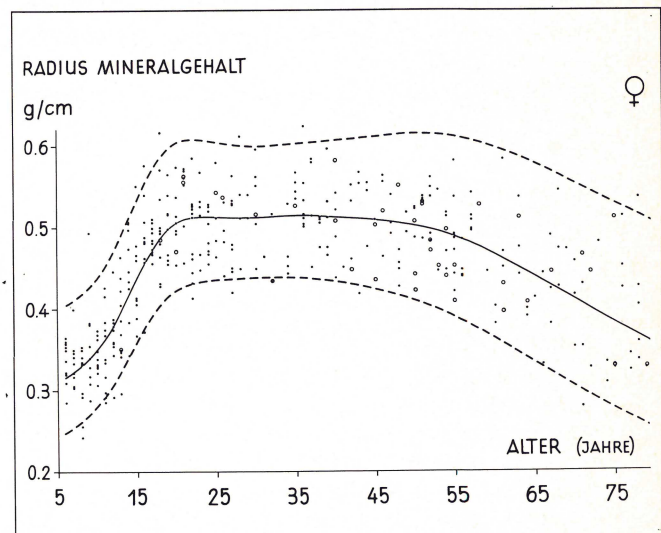


Tabelle 1. Verteilung der Meßwerte um den 50% – Gürtel bei verschiedenen Regressionskurven

	Kurve 1					Kurve 2					Kurve 3					Kurve 4				
	(j = 3)																			
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
unter	23	1	24	23	13	27	7	21	24	20	33	2	22	26	14	25	8	18	21	18
innerh.	44	31	40	43	46	39	27	40	40	36	29	32	43	35	43	39	32	40	41	30
über	8	43	11	9	20	9	41	14	11	21	13	41	10	14	22	11	35	17	13	19
	$\chi^2 = 72.61$					$\chi^2 = 52.30$					$\chi^2 = 68.56$					$\chi^2 = 29.88$				
	SAQ = 3.5755					SAQ = 5.9003					SAQ = 5.0233					SAQ = 6.2870				
	(j = 4)																			
unter	21	7	29	15	15	27	10	26	16	22	27	10	30	17	18	23	8	21	14	17
innerh.	43	36	39	39	47	35	33	38	40	34	31	29	37	35	41	36	37	38	47	31
über	11	32	7	21	17	13	32	11	19	21	17	36	8	23	20	16	30	16	14	19
	$\chi^2 = 38.12$					$\chi^2 = 27.21$					$\chi^2 = 41.23$					$\chi^2 = 20.33$				
	SAQ = 3.2228					SAQ = 5.6728					SAQ = 4.4566					SAQ = 6.0488				

Tabelle 1. Fortsetzung

	Kurve 1					Kurve 2					Kurve 3					Kurve 4				
	(j = 5)																			
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
unter	20	10	27	16	19	27	10	25	21	22	25	11	25	19	21	23	9	19	16	17
innerh.	40	40	39	38	45	33	38	38	33	35	30	34	40	32	43	34	36	38	45	31
über	15	25	9	21	15	15	27	12	21	20	20	30	10	24	15	18	30	18	14	19
	$\chi^2 = 18.71$					$\chi^2 = 18.98$					$\chi^2 = 24.09$					$\chi^2 = 16.84$				
	SAQ = 3.0424					SAQ = 5.5057					SAQ = 4.2506					SAQ = 5.9843				
	(Spline)																			
unter	18	15	23	17	19	18	15	27	20	22	21	15	19	19	21	19	13	21	14	16
innerh.	42	41	39	39	43	45	39	32	36	36	35	36	44	35	39	40	35	34	46	32
über	15	19	13	19	17	12	21	16	19	19	19	24	12	21	19	16	27	20	15	19
	$\chi^2 = 6.135$					$\chi^2 = 10.59$					$\chi^2 = 6.834$					$\chi^2 = 11.10$				
	SAQ = 2.8523					SAQ = 5.0660					SAQ = 3.9327					SAQ = 5.5721				
	I– V: 1. bis 5. Quantil der Altersverteilung																			
	j: Grad des Polynoms																			

angegeben. Die Rechnungen wurden für die Polynomgrade drei, vier und fünf sowie für den kubischen Spline durchgeführt.

Wie man sieht, liegen bei den Splinefunktionen nicht nur die χ^2 -Werte bei ihrem Erwartungswert, sondern die Splinefunktionen weisen auch die kleinsten Abweichungssummen auf.

III. Validierung mittels neuer Meßwerte

Wie aus den Abbildungen 3 und 4 zu ersehen ist, liegen sowohl die neu hinzugenommenen Meßwerte – als Kreise markiert – der 42 Männer als auch die der 42 Frauen innerhalb der 95-Prozentgrenze. Insgesamt über alle Kurven betrachtet, lag der Anteil der Meßwerte, die nicht in die Referenzbereiche fielen, bei 4,8 Prozent und dieser Prozentsatz entspricht somit der Erwartung.

Diskussion

Wie die Ergebnisse in diesem Anwendungsfall zeigen, hat die glättende Splinefunktion gegenüber den Ausgleichspolynomen höherer Ordnung den Vorteil, daß sie nicht nur die Mittelwerte mit ausreichender Genauigkeit schätzt, sondern darüber hinaus auch einen glatten Verlauf nimmt und nicht, wie es den Polynomen höherer Ordnung zukommt, merklich von den zufälligen Schwankungen der Mittelwerte beeinflusst wird.

Dieses Verhalten der Spline ist wohl auch der Grund, daß Splinefunktionen in den letzten Jahren immer häufiger zur Beschreibung nichtlinearer Abhängigkeiten benutzt werden, wobei auch Konfidenzintervalle für die Erwartungswerte abgeleitet werden. Die Benutzung von Splinefunktionen mit Ableitung von Vertrauensgrenzen für Einzelwerte ist uns jedoch aus der Literatur nicht bekannt. Das von uns gewählte, heuristische Verfahren, durch die quadrierten Abweichungen zwischen den beobachteten und mittels Spline geschätzten Werten erneut eine Ausgleichsfunktion zu legen und die Quadratwurzel aus dieser Streufunktion als Standardabweichung der Einzelwerte zu benutzen, erschien uns intuitiv sinnvoll. Die Ergebnisse führten auch unserer Meinung nach auf biologisch brauchbare Grenzen; im jugendlichen Alter ist die Variabilität geringer als in höherem Erwachsenenalter, in dem bedingt durch die Le-

bensweise die Demineralisierung bei einzelnen Personen zu recht unterschiedlichen Zeitpunkten einzutreten scheint. Dementsprechend erhält man auch engere Referenzbereiche in der Jugend und ausgeweitete im höheren Alter. Alternativ hätte man auch Altersklassen bilden und mit den herkömmlichen Verfahren Vertrauensgrenzen berechnen können. Zudem würde man in diesem Fall erheblich mehr Probanden benötigen, um stabile Schätzungen der Variabilität zu erhalten. Der Vorteil der hier vorgestellten Methode liegt darin, daß die Information über die Streuung der Meßwerte benachbarter Altersjahrgänge mit für die Schätzung der Variabilität für ein bestimmtes Alter benutzt wird.

Die versuchte Validierung dieses Vorgehens durch die Hinzunahme von weiteren 42 Männern wie Frauen reicht sicherlich nicht aus; die Verteilung dieser Meßwerte innerhalb der Referenzgürtel spricht jedenfalls nicht gegen dieses vorgeschlagene Verfahren zur Konstruktion von altersabhängigen Referenzbereichen bei nichtlinearen Beziehungen.

Für diagnostische Zwecke würde es natürlich notwendig sein, entsprechende Berechnungen auch an »pathologischen Fällen« vorzunehmen.

Die Tatsache, warum bei der Schätzung der Streufunktion im Gegensatz zur Regression der Weg über die Splinefunktion zu keinem befriedigenden Ergebnis führte, kann von uns nur als Tatsache ohne theoretische Begründung genannt werden.

Literatur

- [1] BÜTTNER, H., Präzisierung von Normalwertbereichen in: Optimierung der Diagnostik, 233–241, Springer, Berlin–Heidelberg–New York 1973
- [2] CAMERON, J. R., SØRENSEN, J. A., Measurement of Bone Mineral in vivo: An Improved Method, *Science* **142**, 230–232, 1963
- [3] CAMERON, J. R., MAZESS, R. B., SØRENSEN, J. A., Precision and Accuracy of Bone Mineral Determination by Direct Photonabsorptiometry, *Invest. Radiol.* **3**, 141–150, 1968
- [4] KUHLENCORDT, F., RINGE, J.-D., KRUSE, H.-P. v. ROTH, A. Bone Mineral Determination of Radius, Ulna and Fingerbones by ^{125}J Photon Absorptiometry on Healthy Persons in: Int. Conf. on Bone Mineral Measurement, Ed.: R. B. MAZESS, Pub.: U. S. Dept. of Health, Education, and Welfare DHEW Publ. No. (NIH), 75–683, 1974
- [5] v. ROTH, A., RINGE, J.-D., KRUSE, H.-P., KUHLENCORDT, F., Bestimmung des Knochenmineralgehaltes durch ^{125}J -Photonenabsorptionsmessung bei Gesunden. *Fortschr. Röntgenstr.* **121**, 597–603, 1974
- [6] RINGE, J.-D., REHPENNING, W., KUHLENCORDT, F., Physiologische Änderung des Mineralgehaltes von Radius und Ulna in Abhängigkeit von Lebensalter und Geschlecht. *Fortschr. Röntgenstr.* **126**, 376–380, 1977
- [7] SCHLENKER, R. A., VON SEGGEN, W. W., The Distribution of Cortical and Trabecular Bone Mass along the Lengths of the Radius and the Ulna and the Implications for in vivo Bone Mass Measurements. *Calcif. Tiss. Res.* **20**, 41–52, 1976
- [8] SPÄTH, H. Algorithmen für elementare Ausgleichmodelle, R. Oldenbourg Verlag, München, 1973
- [9] SPÄTH, H., Spline-Algorithmen zur Konstruktion glatter Kurven und Flächen, R. Oldenbourg Verlag, München 1973

Eingegangen am 9. 10. 80

Anschrift der Verfasser: W. Rehpennig, J. Berger, Universitätskrankenhaus Hamburg-Eppendorf, Abteilung für Mathematik in der Medizin, Martinistr. 52, 2000 Hamburg 20. J.-D. Ringe, Universitätskrankenhaus Hamburg-Eppendorf, Abteilung für Osteologie, Martinistr. 52, 2000 Hamburg 20

Beiträge zur Erfassung der phänotypischen Stabilität

II. Modifikationen des Regressionsansatzes durch Verwendung gewogener Regressionsmodelle sowie Änderungen der Bezugsbasis

M. Hühn

Zusammenfassung

Bei der Analyse von Genotyp-Umwelt-Wechselwirkungen im Rahmen der Entwicklung quantitativer Maßzahlen zur Erfassung der phänotypischen Stabilität wird beim üblichen Regressionsansatz die Leistung x_{ij} des Genotyps i in der Umwelt j in Beziehung gesetzt zur Leistungsfähigkeit u_j der Umwelt j : $x_{ij} = \beta_i u_j + c_i + d_{ij}$ (für jedes i).

Zur Schätzung von u_j werden neben den bisher gebräuchlichen Ansätzen (arithmetischer Mittelwert, arithmetisches Mittel einer Standardgruppe) einige weitere Möglichkeiten vorgeschlagen und diskutiert: Dichtemittel, Zentralwert, Maximalwert und Minimalwert der betreffenden Umwelt j .

Die Schätzung der genotypspezifischen Stabilitätsparameter β_i und $\sum_j d_{ij}^2$ erfolgt hier über einen gewogenen Regressionsansatz

$$» \sum_j d_{ij}^2 w_{ij} = \text{Minimum}«,$$

da den verschiedenen Abweichungen d_{ij} unterschiedliche Gewichte w_{ij} zukommen. Neben dem üblichen ungewichteten Vorgehen ($w_{ij} = 1$) werden als Gewichtungsfaktoren verwendet: reziprokes Interaktionsglied, reziproke Grenzdifferenz, Leistungswert x_{ij} sowie die verschiedenen Schätzwerte für die Leistungsfähigkeit der Umwelt.

Sämtliche Ansätze (Kombinationen verschiedener u_j -Schätzwerte mit verschiedenen Gewichten w_{ij}) werden für ein numerisches Anwendungsbeispiel von 20 Sorten und 10 Orten einer Winterweizen-Wertprüfung des Bundessortenamtes durchgerechnet.

Die Ähnlichkeit bzw. Verschiedenheit der auf diese Weise erhaltenen unterschiedlichen Sortenrangfolgen bezüglich Stabilität läßt sich dann mit Hilfe von Rangkorrelationskoeffizienten vergleichen. Es zeigt sich, daß die verschiedenen Gewichte in sämtlichen untersuchten Fällen zu mehr oder weniger identischen Rangfolgen der Sorten bezüglich Stabilität führen, d. h. die Verwendung dieser gewogenen Regression war – verglichen mit dem üblichen ungewichteten Ansatz – praktisch ohne besonderen Einfluß.

Die verschiedenen verwendeten Bezugsbasen führten dagegen aber zum Teil zu extrem unterschiedlichen Sortenrangfolgen bezüglich Stabilität, wobei jedoch das zugrundegelegte Stabilitätskonzept (maximale Stabilität bei $\beta = 0$ oder $\beta = 1$) eine besondere Rolle spielt.

Für alle Bezugsbasen erweisen sich die Stabilitätsparameter β_i bzw. $|1-\beta_i|$ und $\sum_j d_{ij}^2$ als unkorreliert. Diese Ansätze und Ergebnisse werden eingehend dargestellt und diskutiert.

Summary

Suitable parameters for estimating phenotypic stability quantitatively can be developed by analysing genotype-environment-interactions. In applying the usual regression approach the yield x_{ij} of genotype i in environment j has been related to the yielding ability u_j of the environment j : $x_{ij} = \beta_i u_j + c_i + d_{ij}$ (for each i).

Besides of the common approaches of estimating u_j (arithmetic mean of all genotypes, arithmetic mean of a group of standard genotypes) some further possibilities have been proposed and discussed: mode, median, maximum-yield and minimum-yield of the environment j respectively.

The estimation of the genotype-specific stability-parameters β_i and $\sum_j d_{ij}^2$ here had been derived by a weighted regression approach:

$$» \sum_j d_{ij}^2 w_{ij} = \text{minimum} »,$$

because the different deviations d_{ij} must be assigned different weights w_{ij} . Besides the usual unweighted approach we have used for these weighting-factors: reciprocal interaction term, reciprocal LSD (least significant difference), yield x_{ij} and the different estimates for the yielding ability of the environment. All these numerous possibilities (combinations of different u_j -estimates with different weights w_{ij}) have been computed for a numerical example of 20 varieties and 10 locations from a winter-wheat trial of the German »Bundessortenamt«.

The similarity resp. dissimilarity of the different stability rank-orders of the varieties obtained in these different situations can be compared using Spearman's rank correlation coefficient.

In all u_j - w_{ij} combinations which have been investigated it follows that the different weights results in nearly identical stability rank-orders of the varieties. That means that the weighted regression is of no significant influence for variety-stability compared with the usual unweighted regression approach.

But the different estimates of u_j causes stability rank-orders of the varieties which are extremely different from each other in many situations. But these results are also mainly dependent on the concept of stability (maximum stability for $\beta = 0$ or $\beta = 1$). For each of the different estimates of u_j one obtains stability parameters β_i and $\sum_j d_{ij}^2$ as well as $|1-\beta_i|$ and $\sum_j d_{ij}^2$ which are uncorrelated.

All these different approaches and results are investigated and discussed in detail.

1 Einleitung

Kriterien zur Beurteilung von verschiedenen Individuen und Populationen sind sowohl im pflanzlichen als auch im tierischen Bereich die beiden Aspekte »Leistungsfähigkeit« und »Leistungsstabilität«.

Die Forderung, daß die zu verwendenden Genotypen oder Populationen ihre Leistungsfähigkeit unter möglichst vielen verschiedenen Umweltbedingungen aufrechterhalten sollen, liegt auf der Hand. Mit dem Terminus »Umwelt« werden hierbei sämtliche nichtgenetischen Variationsursachen zusammengefaßt – wie z. B. im pflanzlichen Bereich: Anbauorte, Untersuchungsjahre, Kulturbedingungen u. a.

Die Gegenüberstellung der beiden Begriffe »Leistungsfähigkeit« und »Leistungsstabilität« deutet darauf hin, daß hier bei der Diskussion der »phänotypischen Stabilität« ein Stabilitätskonzept verwendet werden soll, das unabhängig von der Leistungshöhe ist.

Die Stabilitätscharakterisierung, daß eine konstante Leistung unter wechselnden Umweltbedingungen erzielt wird, erweist sich im agronomischen Bereich weniger geeignet als die Forderung, daß die erbrachte Leistung in allen Umwelten in einer bekannten Beziehung zur Leistungsfähigkeit der betreffenden Umwelt steht. Dieses Konzept setzt also voraus, daß es möglich ist, die Leistungsfähigkeit einer Umwelt durch Maßzahlen quantitativ zu kennzeichnen. Die Stabilitätsbetrachtung beruht in solchen Fällen auf einer Untersuchung der Beziehung zwischen der Leistung des Genotyps in einer bestimmten Umwelt und dem entsprechenden Umweltmaß dieser Umwelt.

Im wesentlichen handelt es sich bei den unterschiedlichen Stabilitätsmaßzahlen um verschiedene Verfahren zur Aufteilung der Genotyp-Umwelt-Interaktion. Für die Darstellung und Diskussion dieser verschiedenen Ansätze sei auf einige zusammenfassende Arbeiten verwiesen: WRICKE (1965), SCHNELL (1967), HANSON (1970), GEIDEL (1971), MÜNZER (1972), UTZ (1972), WOLFRAM (1973), FREEMAN (1973), WEILING (1973), HAUFÉ und GEIDEL (1978a und 1978c), WRICKE und WEBER (1980).

2 Problemstellung

Im folgenden wird der Wert des Genotyps i in der Umwelt j mit x_{ij} , $i = 1, 2, \dots, m$ und $j = 1, 2, \dots, n$ bezeichnet. Dabei können diese x_{ij} selbst natürlich wieder Mittelwerte aus verschiedenen Wiederholungen, Parzellen usw. sein. Wir werden wieder durchweg den speziellen Fall im Auge haben, daß den Genotypen »Sorten« und den Umwelten »Standorte« entsprechen, wobei dann x_{ij} z. B. den Ertrag der Sorte i am Standort j bedeuten kann.

Die Maßzahl für die Leistungsfähigkeit der Umwelt j sei u_j – sie werde ohne Fehler gemessen. Die zuvor angesprochene Beziehung zwischen der Leistung eines Genotyps in einer bestimmten Umwelt und dem entsprechenden Umweltmaß dieser Umwelt betrifft also – für jedes i – die Relation zwischen den x_{ij} und den u_j . Die einfachste Abhängigkeit wäre eine lineare Beziehung – und so lautet also der entsprechende lineare Regressionsansatz:

$$x_{ij} = \beta_i u_j + c_i + d_{ij} \quad (\text{für jedes } i)$$

Dabei bezeichnet β_i den Regressionskoeffizienten, c_i die additive Konstante der Regressionsgleichung und d_{ij} die Abweichung von der Regression. Die Parameter β_i und c_i schätzt man üblicherweise über die Bedingung: $\sum_j d_{ij}^2 = \text{Minimum}$.

Bei der Anwendung dieses Modells erhebt sich natürlich sofort die Frage: Wie kommt man zur »wahren« Leistungsfähigkeit u_j der Umwelt j bzw. wie kann man u_j schätzen? Dies ist ein Hauptproblem des beschriebenen Regressionsansatzes.

Eine Schätzung von u_j anhand der Ausgangsdaten x_{ij} liefert kein unabhängiges Umweltmaß u_j und führt damit zu den bekannten Schwierigkeiten bei der statistischen Auswertung. Zur Vermeidung dieser Probleme wurden verschiedene Lösungsvorschläge gemacht, wie z. B.:

- 1) Aufteilung des Materials und Verwendung des einen Teiles zur u_j -Schätzung und des anderen Teiles zur Regressionsauswertung: FREEMAN und PERKINS (1971); FRIPP (1972) u. a.
- 2) Charakterisierung der Umwelteleistungsfähigkeit u_j anhand von physischen Umweltvariablen wie Temperatursummen, tägliche Niederschläge, Luftfeuchtigkeit, physikalische Bodendaten usw.: HARDWICK und WOOD (1972), FATUNLA und FREY (1976); SEIF und PETERSON (1978); HAUFÉ und GEIDEL (1978b und 1978c) u. a.

Für eine nähere Diskussion dieser Ansätze sei auf die angegebene Literatur verwiesen.

Da in den meisten Fällen jedoch kein solches unabhängiges Umweltmaß u_j zur Verfügung steht, wird u_j unter Heranziehung der Ausgangsdaten x_{ij} geschätzt – und man verwendet hierbei dann meistens den arithmetischen Umweltmittelwert

$$\bar{x}_{.j} = \frac{1}{m} \sum_{i=1}^m x_{ij}$$

als Schätzwert für u_j . Die Maßzahl »arithmetischer Mittelwert« und damit dann auch die Schätzung $\hat{u}_j = \bar{x}_{.j}$ weist jedoch bekanntermaßen einige Nachteile auf. Hier bieten sich verschiedene andere, möglicherweise bessere Alternativen an, die im folgenden Kapitel 3 im Unterabschnitt »Wahl der Bezugsbasis« untersucht werden sollen.

Eine zweite Modifikation des beschriebenen Regressionsansatzes betrifft die Parameterschätzung aus der Bedingung $\sum_j d_{ij}^2 = \text{Minimum}$.

Bei diesem Vorgehen gehen alle d_{ij} -Glieder völlig gleichwertig in die Berechnungen ein. Auch diese Annahme dürfte eine etwas unrealistische Voraussetzung darstellen, da – für ein bestimmtes i – den einzelnen d_{ij} d. h. dann also den verschiedenen Umwelten j , ein unterschiedliches »Gewicht« zuzuschreiben ist. Solche unterschiedlichen Gewichte der verwendeten Umwelten können z. B. durch die unterschiedliche Genauigkeit begründet sein, mit der die Leistungsfähigkeit dieser Umwelten geschätzt werden kann.

Bezeichnet man das Gewicht von d_{ij} mit w_{ij} , so lautet nun die Bedingung zur Schätzung der Parameter:

$$\sum_j d_{ij}^2 w_{ij} = \text{Minimum}.$$

Dieser Ansatz führt auf die folgenden beiden Gleichungen zur Schätzung der Parameter β_i und c_i :

$$\hat{\beta}_i \cdot \left(\sum_{j=1}^n u_j w_{ij} \right) + \hat{c}_i \cdot \left(\sum_{j=1}^n w_{ij} \right) = \sum_{j=1}^n x_{ij} w_{ij} \quad (2)$$

$$\hat{\beta}_i \cdot \left(\sum_{j=1}^n u_j^2 w_{ij} \right) + \hat{c}_i \cdot \left(\sum_{j=1}^n u_j w_{ij} \right) = \sum_{j=1}^n u_j x_{ij} w_{ij} \quad (3)$$

Für jedes i wird die Variabilität um die Regression gemessen durch den Ausdruck $\sum_j d_{ij}^2$, der geschätzt wird durch

$$\sum_{j=1}^n (x_{ij} - \hat{\beta}_i \hat{u}_j - \hat{c}_i)^2 = \sum_{j=1}^n \hat{d}_{ij}^2 \quad (4)$$

Für die zu verwendenden Gewichtungsfaktoren w_{ij} bieten sich eine ganze Reihe von Möglichkeiten an, die im folgenden Kapitel 3 im Unterabschnitt »Festlegung der Gewichte« diskutiert werden sollen.

Im Kapitel 4 werden dann sämtliche Ansätze an einem numerischen Anwendungsbeispiel demonstriert.

3 Überlegungen zur Wahl der Bezugsbasis u_j sowie zur Festlegung der Gewichte w_{ij}

3.1 Festlegung der Gewichte

Die Verwendung gewogener Ansätze bei der Analyse von Genotyp-Umwelt-Wechselwirkungen wurde in der Literatur zwar verschiedentlich erwähnt (siehe z. B. RUNDFELDT 1980), bisher jedoch nicht eingehender angewendet und untersucht. Zumindest dürfte dieses für solche gewichteten Auswertungen im Rahmen der Berechnung der genotypspezifischen Regressionsen der x_{ij} auf die u_j (für jedes i) zutreffen.

Die Hauptschwierigkeit liegt hierbei natürlich in einer sinnvollen Auswahl der Gewichte w_{ij} .

Naheliegender wäre natürlich eine Verwendung der Genotyp-Umwelt-Interaktionen selbst bei dieser Gewichtung, d. h. man zieht für w_{ij} das Interaktionsglied des Genotyps i in der Umwelt j heran. Dies kann z. B. in der folgenden Weise geschehen: Da für die Stabilitätsbetrachtung nur die Größe des Interaktions termes, jedoch nicht sein Vorzeichen interessiert, wird der absolute Betrag des Interaktionsgliedes eingeführt:

$$a_{ij} = |x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}| \quad (5)$$

Dabei ist $\bar{x}_{i.} = \frac{1}{n} \sum_j x_{ij}$ der Mittelwert des Genotyps i , $\bar{x}_{.j} = \frac{1}{m} \sum_i x_{ij}$ der Mittelwert der Umwelt j und $\bar{x}_{..} = \frac{1}{mn} \sum_{ij} x_{ij}$ der Gesamtmittelwert.

Die Verwendung von $\frac{1}{a_{ij}}$ als Gewicht w_{ij} hätte zwei besondere Nachteile: Dieser Ausdruck wäre für $a_{ij} = 0$ (ein Fall, der ja durchaus vorkommt) nicht definiert und zum anderen würde hierbei sehr kleinen a_{ij} -Werten (wegen $\frac{1}{a_{ij}} \rightarrow \infty$ für $a_{ij} \rightarrow 0$) ein unrealistisch hohes Gewicht zukommen. Diese Schwierigkeit läßt sich z. B. vermeiden, indem man vom Mittelwert $\bar{a}_{i.} = \frac{1}{n} \sum_{j=1}^n a_{ij}$ als Bezugspunkt ausgeht, d. h. für $f_{ij} = \bar{a}_{i.} + a_{ij}$ wird gesetzt:

$$w_{ij} = f_{ij}^{-1} = \frac{1}{\bar{a}_{i.} + a_{ij}} \quad (6)$$

Eine weitere auf der Hand liegende Möglichkeit zur Festlegung der Gewichte w_{ij} besteht in einer Heranziehung des Versuchsfehlers der Umwelt j , d. h. Umwelten mit hohen Versuchsfehlern soll ein weniger starkes Gewicht zukommen als Umwelten, deren Versuchsfehler geringer ist. Dieses leistet z. B. die folgende w_{ij} -Definition:

$$w_{ij} = \frac{1}{(GD)_j} \quad \text{für jedes } i \quad (7)$$

Hierbei bezeichnet $(GD)_j$ die Grenzdifferenz der Umwelt j zu einer bestimmten vorgegebenen Irrtumswahrscheinlichkeit.

Neben den beiden Situationen (6) und (7) wurden zusätzlich noch die folgenden Fälle untersucht:

$$w_{ij} = 1 \quad \text{für alle } i \text{ und alle } j \quad (8)$$

Dies wäre der übliche ungewichtete Ansatz, d. h. der Fall gleicher Gewichte.

$$w_{ij} = x_{ij} \quad (9)$$

Jede Abweichung d_{ij} wird bei dieser w_{ij} -Definition also gewogen mit dem jeweiligen Merkmalswert x_{ij} .

$$w_{ij} = u_j \quad \text{für jedes } i \quad (10)$$

Bei diesem Ansatz wird als Gewicht w_{ij} die Leistungsfähigkeit der betreffenden Umwelt j herangezogen. Hierbei können für u_j nun die im nachfolgenden Abschnitt 3.2. beschriebenen verschiedenen Möglichkeiten zur Schätzung der Umwelt-Leistungsfähigkeit verwendet werden. Lediglich bei (6) und (9) werden sowohl genotyp- als auch umweltspezifische Gewichte gewählt, während in allen anderen Fällen die Gewichte w_{ij} für alle Genotypen i übereinstimmen und sie nur von der Umwelt j abhängen.

3.2 Wahl der Bezugsbasis

Ein schwerwiegender Nachteil des arithmetischen Mittelwertes und damit auch der Schätzung $\hat{u}_j = \bar{x}_{.j}$ ist die starke Sortimentsabhängigkeit und besonders die große Empfindlichkeit gegenüber Extremwerten. Unter der Voraussetzung eines genügend umfangreichen und repräsentativen Sortiments dürften daher andere Mittelwertmaße oftmals geeigneter sein. Hierfür kommen zunächst das »Dichtemittel« sowie der »Zentralwert« in Frage.

Das Dichtemittel ist der häufigste Wert der Ausgangsdaten, während der Zentralwert oder Median den Wert bezeichnet, der die der Größe nach geordneten Ausgangsdaten in zwei gleiche Hälften teilt. Diese beiden Mittelwertmaße weisen die Empfindlichkeit des arithmetischen Mittels gegenüber Extremwerten nicht auf. Außerdem charakterisieren sie in einer inhaltlich begründeteren Weise die »mittlere Tendenz« der Daten, während der wegen seiner günstigeren statistischen Eigenschaften meistens angewendete arithmetische Mittelwert oft nur eine mehr oder weniger rein rechnerische Konstruktion darstellt.

Wir untersuchen daher die folgenden Fälle:

$$\hat{u}_j = \text{Dichtemittel der } \{x_{ij} | i = 1, 2, \dots, m; j = \text{fest}\} \quad (11)$$

$$\hat{u}_j = \text{Zentralwert der } \{x_{ij} | i = 1, 2, \dots, m; j = \text{fest}\} \quad (12)$$

$$\hat{u}_j = \bar{x}_{.j} = \text{Arithmetisches Mittel der } \{x_{ij} | i = 1, 2, \dots, m; j = \text{fest}\} \quad (13)$$

Sieht man das vorliegende Sortiment als ein bestimmtes, zur Verfügung stehendes genetisches Potential an, dessen phänotypische Ausprägung unter verschiedenen Umweltbedingungen j betrachtet wird, so wäre unter agronomischen Gesichtspunkten die Leistungsfähigkeit einer Umwelt j auch durch den Maximalertrag in dieser Umwelt j zu charakterisieren, denn unter Ausnutzung des vorgegebenen genetischen Potentials hat die Umwelt j diese Leistung ja auch tatsächlich realisiert. Die Heranziehung der Maximalerträge bei der Erfassung der phänotypischen Stabilität wurde bereits verschiedentlich vorgeschlagen (siehe z. B.: JENSEN (1976); FISCHBECK (1976); HAUFEL und GEIDEL (1978a)). Dies geschah jedoch meist in der Weise,

daß für jeden Genotyp der mittlere Abstand vom jeweiligen Höchstertag in den verschiedenen Umwelten zur Stabilitätscharakterisierung verwendet wurde. Bisher noch nicht eingehender untersucht wurde jedoch die Heranziehung des Maximalertrages als Schätzwert für die Leistungsfähigkeit u_j der Umwelt j bei Berechnung der genotyp-spezifischen Regression der x_{ij} auf die u_j :

$$\hat{u}_j = \text{Maximum der } \{x_{ij} | i = 1, 2, \dots, m; j = \text{fest}\} \quad (14)$$

Von einem gewissen Interesse im Rahmen dieser Überlegungen zur Schätzung der Leistungsfähigkeit der Umwelt j könnte sogar auch die Verwendung des Minimalertrages sein:

$$\hat{u}_j = \text{Minimum der } \{x_{ij} | i = 1, 2, \dots, m; j = \text{fest}\} \quad (15)$$

Denn: Was eine Umwelt – bei vorgegebenem genetischen Potential – mindestens zu leisten imstande ist, charakterisiert ja auch in einer bestimmten Weise die Leistungsfähigkeit dieser Umwelt. Höchste Leistungsfähigkeit entspricht hierbei also einem maximalen Minimalertrag.

In größeren Versuchsserien, wie z. B. den Wertprüfungen des Bundessortenamtes, ist die Verwendung einer Standardortengruppe gebräuchlich. Dies sind Sorten, die in allen Versuchen als Vergleichsgruppe mitgeführt werden. Die Heranziehung des Mittelwertes einer Standardgruppe zur Schätzung der Leistungsfähigkeit der Umwelt ist daher häufig diskutiert worden (siehe z. B.: LEIN und HAUF (1967); FREEMAN und PERKINS (1971); HAUF und GEIDEL (1978c)). Auch diese Schätzmöglichkeit für u_j soll bei den nachfolgenden Auswertungen mit einbezogen werden:

$$\hat{u}_j = \text{Mittelwert der Standardgruppe in der Umwelt } j \quad (16)$$

Für die beiden Einflußgrößen »Gewichte« und »Bezugsbasen« stehen daher die folgenden Möglichkeiten zur Diskussion:

Gewichte w_{ij}

- $w_{ij} = f_{ij}^{-1}$: Reziprokes Interaktionsglied
- $w_{ij} = 1$: Gleiche Gewichte, d. h. ungewichtet
- $w_{ij} = (GD)_j^{-1}$: Reziproke Grenzdifferenz
- $w_{ij} = x_{ij}$: Leistungswerte
- $w_{ij} = u_j$ mit \hat{u}_j = Arithmetischer Mittelwert der Umwelt j
- $w_{ij} = u_j$ mit \hat{u}_j = Dichtemittel der Umwelt j
- $w_{ij} = u_j$ mit \hat{u}_j = Zentralwert der Umwelt j
- $w_{ij} = u_j$ mit \hat{u}_j = Maximalwert der Umwelt j
- $w_{ij} = u_j$ mit \hat{u}_j = Minimalwert der Umwelt j
- $w_{ij} = u_j$ mit \hat{u}_j = Arithmetisches Mittel der Standardgruppe in der Umwelt j

Bezugsbasis u_j

- \hat{u}_j = Arithmetischer Mittelwert der Umwelt j
- \hat{u}_j = Dichtemittel der Umwelt j
- \hat{u}_j = Zentralwert der Umwelt j
- \hat{u}_j = Maximalwert der Umwelt j
- \hat{u}_j = Minimalwert der Umwelt j
- \hat{u}_j = Arithmetisches Mittel der Standardgruppe in der Umwelt j

Für die Kombinationen eines jeden w_{ij} mit einem jeden u_j (insgesamt also für 60 verschiedene Fälle) werden im folgenden Kapitel 4 die Regressionsberechnungen für ein numerisches Anwendungsbeispiel durchgeführt und diskutiert.

4 Anwendungen auf ein numerisches Beispiel

Bei dem verwendeten Zahlenmaterial handelt es sich wieder um die gleichen Daten, die auch in der ersten Arbeit dieser Veröffentlichungsreihe »Beiträge zur Erfassung der phänotypischen Stabilität« (HÜHN (1979)) als Anwendungsbeispiel ausgewertet wurden. Es ist ein Auszug aus den Ergebnissen der Wertprüfung II mit Winterweizen (Sortiment A) des Bundessortenamtes für das Erntejahr 1974. Untersucht wird im folgenden nur das Merkmal »Kornertag« (in dt/ha) bei 86% Trockensubstanz für die Durchschnittsergebnisse der beiden untersuchten Stufen (Stufe 1 = normale Wirtschaftsweise; Stufe 2 = intensivere Wirtschaftsweise). Der vollständige Datensatz von 42 Sorten und 16 Orten wird im Rahmen einer späteren Publikation verrechnet werden; hier wurden als Anwendungsbeispiel 20 Sorten und 10 Orte ausgewählt. Für die Sorten und die Orte sind die Original-Kurzbezeichnungen des Bundessortenamtes beibehalten worden. Die Kornträge der 20 Sorten an den 10 Orten – d. h. die Ausgangsdaten x_{ij} für alle folgenden Auswertungen – sind in der Tabelle 1 zusammengefaßt. Die Standardortengruppe besteht aus den drei Sorten JUBILAR, DIPLOMAT und CARIBO.

Für einige ausgewählte Fälle von Gewichten w_{ij} und Bezugsbasen u_j sind die Ergebnisse der Regressionsberechnungen (β_i und $\sum_j d_{ij}^2$) in der Tab. 2 wiedergegeben. Für eine bestimmte ausgewählte Bezugsbasis führen die verschiedenen Gewichte (unter Einschluß des Falles »ungewichtet«) zu praktisch gleichen Ergebnissen – und zwar sowohl bei den Regressionskoeffizienten als auch bei den Summen der Abweichungsquadrate um die Regression. Für verschiedene Bezugsbasen ergeben sich jedoch mehr oder weniger unterschiedliche Resultate für β_i und $\sum_j d_{ij}^2$. Diese für zwei Bezugsbasen und drei Gewichte geltenden Ergebnisse der Tab. 2 besagen also, daß für den vorliegenden Datensatz die Verwendung von Gewichten nicht erforderlich ist, während die Auswahl der Bezugsbasis jedoch eine bedeutendere Rolle spielt. Diese Schlußfolgerung erweist sich auch bei der Untersuchung aller anderen Bezugsbasen und aller anderen Gewichte als zutreffend. Eine etwas genauere Betrachtung dieser Ergebnisse soll im folgenden für die beiden Stabilitätsparameter β_i und $\sum_j d_{ij}^2$ getrennt durchgeführt werden.

4.1 Ergebnisse für den Stabilitätsparameter $\sum_j d_{ij}^2$

Bei jeder der verwendeten sechs Bezugsbasen führt die Anwendung der zehn verschiedenen Gewichte zu jeweils zehn praktisch identischen Sortenrangfolgen bezüglich $\sum_j d_{ij}^2$. Die entsprechenden Spearman'schen-Rangkorrelationskoeffizienten zwischen diesen 10 Rangfolgen liegen sämtlich über 0,99 – und dieses Ergebnis trifft für jede der sechs Bezugsbasen zu. Nach erfolgter Wahl der Bezugsbasis ist also bei den Regressionsrechnungen – in bezug auf $\sum_j d_{ij}^2$ – die Verwendung von Gewichten nicht erforderlich.

Die nächste interessierende Frage wäre: Wie unterscheiden sich die Stabilitätsrangfolgen der Sorten bezüglich $\sum_j d_{ij}^2$ für die verschiedenen verwendeten Bezugsbasen? Da die Gewichte praktisch ohne Bedeutung sind, wird der folgenden Betrachtung der Bezugsbasen jeweils der Fall »gleiche Gewichte« zugrundegelegt. Man erhält die Ergebnisse der Tabelle 3. Die Rangkorrelationskoeffizienten zwischen arithmetischem Mittelwert, Dichtemittel, Zentralwert und – mit einer gewissen Abstufung – auch dem arithmetischen Mittel der Standardgruppe erweisen sich untereinander jeweils als relativ hoch (0,82 bis

Tabelle 1. Kornerträge (dt/ha) der 20 Sorten an 10 Orten sowie verschiedene standortspezifische Hilfsgrößen wie: Grenzdifferenz, arithmetisches Mittel, Zentralwert, Dichtemittel, Maximal- und Minimalwert sowie arithmetisches Mittel der Standardgruppe

Orte Sorten	WUHG	N SCHON	RETHM	OSTIN	DIKOP	WENG	LAD	HOHNH	KLOST	GISEN
JUBILAR	76,3	79,0	64,4	65,2	68,3	51,7	60,4	75,8	53,0	70,4
DIPLOMAT	78,7	83,5	63,1	57,4	67,9	54,5	58,7	81,0	51,1	75,1
CARIBO	78,2	83,9	67,1	74,0	65,3	52,6	61,0	83,8	52,7	71,8
CBC 710	79,4	87,8	69,6	71,9	77,2	53,3	66,5	76,6	47,9	80,1
STRU 721	88,1	88,1	68,5	76,8	71,1	49,4	63,4	80,2	53,5	72,2
ACK 712	78,2	78,9	65,3	66,8	70,8	47,4	62,2	73,6	37,4	70,5
BEUN 781	91,4	90,9	71,3	69,8	80,3	49,7	70,7	85,8	53,1	85,7
BREU 737	68,2	74,0	60,5	66,0	69,6	46,1	59,4	76,7	45,4	67,0
BRND 758	75,1	79,1	64,6	52,8	74,9	52,4	53,5	87,9	61,0	71,6
DOER 750	93,8	80,4	60,8	70,3	68,6	51,3	63,5	78,8	49,9	68,6
FIRL 777	77,7	80,4	65,4	68,6	64,5	51,8	56,1	79,5	46,0	76,0
BRND 759	88,3	72,3	62,8	57,5	71,0	49,3	57,8	79,7	59,6	77,2
FIRL 779	71,4	76,9	63,1	60,9	65,5	55,8	54,3	81,1	48,6	65,5
FIRL 780	79,4	82,0	68,9	73,8	69,1	57,1	59,8	80,2	53,1	72,9
FRAH 743	89,4	77,4	63,3	69,8	75,2	47,0	62,6	82,8	62,6	78,8
LOCH 744	84,7	76,5	72,7	71,6	73,4	44,7	65,0	82,3	59,2	74,5
LOCH 745	82,9	81,1	71,1	67,7	68,6	50,1	65,5	81,6	60,7	77,9
RUEM 711	81,9	74,8	67,4	64,2	72,9	56,6	59,3	78,0	58,0	77,5
PEM 2	85,4	83,6	71,5	71,5	70,7	55,6	59,7	87,1	51,6	80,5
PEM 3	86,9	86,3	73,5	74,2	73,4	52,6	62,5	80,7	52,3	79,1
Grenzdif.	7,19	4,00	4,61	5,59	5,21	3,70	5,10	5,59	4,10	4,50
arithm. Mittel	81,8	80,8	66,7	67,5	70,9	51,4	61,0	80,7	52,8	74,6
Zentralwert	81,8	80,2	66,5	68,5	70,9	52,0	60,5	80,3	53,5	74,4
Dichtemittel	80,7	77,9	64,5	68,2	71,0	52,2	59,1	80,2	53,8	73,9
Maximalwert	93,8	90,9	73,5	76,8	80,3	57,1	70,7	87,9	62,6	85,7
Minimalwert	68,2	72,3	60,5	52,8	64,5	44,7	53,5	73,6	37,4	65,5
arithm. Mittelwert d. Stand.	77,7	82,1	64,9	65,5	67,2	52,9	60,0	80,2	52,3	72,4

Tabelle 2. Ergebnisse der Regressionsberechnungen (β_i und $\sum_j d_{ij}^2$) für einige ausgewählte Fälle von Gewichten w_{ij} und Bezugsbasen u_j .

Bezugsbasis Gewicht	Arithm. Mittel ungewichtet	Arithm. Mittel rez. Grenzdif.	Arithm. Mittel rez. Interakt.	Maximalwert ungewichtet	Maximalwert rez. Grenzdif.	Maximalwert rez. Interakt.
Sorten	$\beta_i \quad \sum_j d_{ij}^2$	$\beta_i \quad \sum_j d_{ij}^2$	$\beta_i \quad \sum_j d_{ij}^2$	$\beta_i \quad \sum_j d_{ij}^2$	$\beta_i \quad \sum_j d_{ij}^2$	$\beta_i \quad \sum_j d_{ij}^2$
JUBILAR	0,84 8,89	0,85 9,11	0,86 9,56	0,76 25,68	0,77 25,95	0,78 26,61
DIPLOMAT	1,00 129,40	1,00 130,92	1,01 129,47	0,90 178,22	0,89 179,90	0,93 179,81
CARIBO	0,97 113,99	0,99 114,34	0,99 114,21	0,87 174,97	0,88 175,44	0,89 175,71
CBC 710	1,04 176,93	1,09 180,80	1,03 179,24	0,94 192,73	0,99 196,44	0,91 198,73
STRU 721	1,12 128,47	1,14 128,77	1,10 129,56	1,03 123,00	1,04 123,12	1,03 123,34
ACK 712	1,12 187,36	1,15 188,30	1,03 199,41	1,01 238,14	1,03 238,76	0,89 260,92
BEUN 781	1,29 94,18	1,32 94,86	1,24 97,14	1,19 65,08	1,21 65,69	1,17 65,89
BREU 737	0,90 115,39	0,93 116,82	0,96 119,60	0,80 165,17	0,84 166,88	0,86 170,88
BRND 758	0,91 437,77	0,89 439,50	0,91 437,85	0,80 510,04	0,79 511,44	0,79 511,03
DOER 750	1,11 231,61	1,06 236,32	1,02 243,01	1,03 200,75	0,98 206,14	0,94 212,24
FIRL 777	1,06 81,68	1,06 82,06	1,04 82,35	0,95 145,74	0,95 146,17	0,92 148,41
BRND 759	0,96 280,81	0,91 285,25	1,01 282,94	0,90 235,74	0,85 240,20	0,94 238,49
FIRL 779	0,84 141,08	0,83 141,84	0,92 150,16	0,72 239,09	0,71 239,68	0,84 258,77
FIRL 780	0,87 67,03	0,87 67,11	0,92 70,79	0,78 118,33	0,78 118,42	0,85 126,84
FRAH 743	1,02 201,15	1,00 203,30	1,04 201,44	0,97 125,11	0,95 126,75	1,01 127,25
LOCH 744	0,98 162,90	0,98 164,48	0,94 167,78	0,90 149,46	0,91 150,69	0,88 153,53
LOCH 745	0,91 75,92	0,91 76,02	0,93 76,59	0,84 64,92	0,85 65,06	0,85 65,06
RUEM 711	0,79 72,21	0,76 73,01	0,84 76,50	0,72 76,59	0,69 77,46	0,76 79,03
PEM 2	1,12 57,03	1,11 57,23	1,07 60,20	1,00 134,37	0,99 134,59	0,93 140,83
PEM 3	1,10 55,03	1,12 55,24	1,11 55,07	1,00 78,01	1,01 78,16	0,99 78,31

Tabelle 3. Rangkorrelationskoeffizienten zwischen den Sortenrangfolgen bezüglich $\sum_j d_{ij}^2$ für die sechs untersuchten Bezugsbasen (* bzw. ** = Signifikanz bei einer Irrtumswahrscheinlichkeit von 5 % bzw. 1 %)

	Arithm. Mittelw.	Maximalwert	Minimalwert	Dichtemittel	Zentralwert	Arithm. Mitt. d. Stand.
Arithm. Mittelw.	–	0,79**	0,59**	0,93**	0,99**	0,83**
Maximalwert	0,79**	–	0,19	0,74**	0,77**	0,46*
Minimalwert	0,59**	0,19	–	0,59**	0,55*	0,73**
Dichtemittel	0,93**	0,74**	0,59**	–	0,95**	0,82**
Zentralwert	0,99**	0,77**	0,55*	0,95**	–	0,85**
Arithm. Mitt. d. Stand.	0,83**	0,46*	0,73**	0,82**	0,85**	–

0,99), während die Rangkorrelationskoeffizienten mit dem Minimalwert und auch mit dem Maximalwert deutlich geringere Werte aufweisen (0,19 – 0,79). Hierbei ergibt sich mit 0,19 der mit Abstand niedrigste Rangkorrelationskoeffizient für den Vergleich von Maximal- und Minimalwert.

4.2 Ergebnisse für den Stabilitätsparameter β_i

Da auch hier wieder Stabilitätsrangfolgen der Sorten bezüglich β_i herangezogen werden sollen, muß zunächst festgelegt werden, was man unter »maximaler Stabilität« verstehen will. Wir betrachten zuerst den Fall, daß eine Sorte mit $\beta = 0$ maximal stabil genannt werden soll. Danach folgt die Diskussion von maximaler Stabilität im Sinne von $\beta = 1$.

4.2.1 Ergebnisse für β_i bei Betrachtung von $\beta = 0$ als maximale Stabilität

Wieder führt bei jeder der verwendeten sechs Bezugsbasen die Anwendung der zehn verschiedenen Gewichte zu jeweils zehn mehr oder weniger identischen Sortenrangfolgen bezüglich β_i . Von den entsprechenden Rangkorrelationskoeffizienten zwischen diesen 10 Rangfolgen (je Bezugsbasis 45 Vergleiche und somit insgesamt $6 \times 45 = 270$ Rangkorrelationskoeffizienten) liegen fast 92 % über 0,90. Der kleinste überhaupt auftretende Wert ist 0,82. Nach erfolgter Wahl der Bezugsbasis ist bei den Regressionsrechnungen des vorliegenden Datensatzes – in bezug auf β_i – die Verwendung von Gewichten nicht erforderlich.

Für den Vergleich der Stabilitätsrangfolgen der Sorten bezüglich β_i für die verschiedenen Bezugsbasen wird – da die Gewichte praktisch ohne Bedeutung sind – wieder der Fall »ungewichtet« zugrundegelegt. Man erhält die Rangkorrelationskoeffizienten der linken unteren Dreiecksmatrix der Tabelle 4.

Sie erweisen sich fast durchweg als extrem hoch. Der kleinste auftretende Wert ergibt sich mit 0,82 für den Vergleich zwischen Maximal- und Minimalwert. Dieses Ergebnis besagt also, daß in dem vorliegenden Falle nicht nur die verschiedenen Gewichte, sondern auch die unterschiedlichen Bezugsbasen zu mehr oder weniger identischen Sortenrangfolgen führen.

4.2.2 Ergebnisse für β_i bei Betrachtung von $\beta = 1$ als maximale Stabilität

Von einigen wenigen Ausnahmen abgesehen führt – bei jeder der verwendeten sechs Bezugsbasen – die Anwendung der zehn verschiedenen Gewichte auch hier wieder zu weitgehend ähnlichen Sortenrangfolgen bezüglich β_i . Von den entsprechenden Rangkorrelationskoeffizienten zwischen diesen 10 Rangfolgen liegen annähernd 60 % über 0,90 und fast 90 % über 0,70. Die erwähnten Ausnahmen betreffen die »reziproke Grenzdifferenz« sowie das »reziproke Interaktionsglied« mit teilweise etwas niedrigeren Werten. In allen anderen Fällen erweisen sich die Rangkorrelationskoeffizienten als besonders hoch.

Für den Vergleich der Stabilitätsrangfolgen der Sorten bezüglich β_i für die verschiedenen Bezugsbasen wird auch hier wieder der Fall »ungewichtet« zugrundegelegt. Man erhält die Werte der rechten oberen Dreiecksmatrix der Tabelle 4. Die Rangkorrelationskoeffizienten zwischen arithmetischem Mittelwert, Dichtemittel, Zentralwert und dem arithmetischen Mittel der Standardgruppe erweisen sich untereinander jeweils als relativ hoch (0,72 – 0,91), während die meisten übrigen Werte ein negatives Vorzeichen aufweisen, wobei jedoch die absoluten numerischen Werte mehr oder weniger klein ausfallen (–0,29 – 0,35). Eine Ausnahme bildet hierbei lediglich der Vergleich von Maximal- und Minimalwert mit einem Rangkorrelationskoeffizienten von 0,79 (siehe Tab. 4).

Tabelle 4. Rangkorrelationskoeffizienten zwischen den Sortenrangfolgen bezüglich β_i für die sechs untersuchten Bezugsbasen (rechte obere Dreiecksmatrix: Werte für den Fall einer maximalen Stabilität bei $\beta = 1$; linke untere Dreiecksmatrix: Werte für den Fall einer maximalen Stabilität bei $\beta = 0$) (* bzw. ** = Signifikanz bei einer Irrtumswahrscheinlichkeit von 5 % bzw. 1 %)

	Arithm. Mittelw.	Maximalwert	Minimalwert	Dichtemittel	Zentralwert	Arithm. Mitt. d. Stand.
Arithm. Mittelwert	–	0,35	0,27	0,72**	0,89**	0,76**
Maximalwert	0,98**	–	0,79**	–0,29	0,02	–0,10
Minimalwert	0,90**	0,82**	–	–0,22	–0,06	–0,22
Dichtemittel	0,98**	0,97**	0,88**	–	0,91**	0,89**
Zentralwert	1,00**	0,97**	0,88**	0,99**	–	0,91**
Arithm. Mitt. d. Stand.	0,98**	0,92**	0,94**	0,97**	0,98**	–

4.3 Vergleich der Ergebnisse für β_i und $\sum_j d_{ij}^2$

Von besonderem Interesse ist die Frage des Zusammenhanges zwischen den β_i - und $\sum_j d_{ij}^2$ -Parametern. Solche Untersuchungen wurden in der Literatur verschiedentlich aufgegriffen und ergaben bei sehr verschiedenen Kulturpflanzen jeweils eine unabhängige Vererbung von β_i und $\sum_j d_{ij}^2$: PERKINS und JINKS (1968) bei Tabak, EBERHART und RUSSELL (1969) bei Mais, BUSCH, HAMMOND und FROBERG (1976) bei Weizen und DHILLON und SINGH (1977) wiederum bei Mais.

Als Beitrag zu dieser Problematik kann im folgenden zwar keine genauere Analyse von abhängiger oder unabhängiger Vererbung, sondern lediglich eine Berechnung der Korrelation zwischen diesen beiden Parametern durchgeführt werden. Bei der Betrachtung der verschiedenen Bezugsbasen wird aufgrund der vorherigen Ergebnisse wieder jeweils nur der Fall »ungewichtet« zugrundegelegt. Berechnet werden die üblichen Korrelationskoeffizienten zwischen β_i und $\sum_j d_{ij}^2$ bzw.

$|1-\beta_i|$ und $\sum_j d_{ij}^2$ sowie die entsprechenden Rangkorrelationskoeffizienten.

Tabelle 5. Korrelationskoeffizienten sowie Rangkorrelationskoeffizienten zwischen den Parametern β_i und $\sum_j d_{ij}^2$ bzw. $|1-\beta_i|$ und $\sum_j d_{ij}^2$ für die verschiedenen Bezugsbasen (jeweils für den Fall »gleicher Gewichte«)

Bezugsbasis	Rangkorrelation $ 1-\beta_i , \sum_j d_{ij}^2$ ($\beta_i, \sum_j d_{ij}^2$)		Korrelation $ 1-\beta_i , \sum_j d_{ij}^2$ ($\beta_i, \sum_j d_{ij}^2$)	
Arithm. Mittelw.	-0,44	+0,15	-0,32	+0,02
Maximalwert	-0,14	+0,01	-0,01	-0,15
Minimalwert	-0,16	+0,18	-0,10	+0,01
Arithm. Mitt. d. Stand.	-0,33	+0,16	-0,27	+0,08
Zentralwert	-0,35	+0,14	-0,29	+0,02
Dichtemittel	-0,30	+0,33	-0,22	+0,19

Die Ergebnisse enthält die Tabelle 5. Für alle Bezugsbasen sind die Rangkorrelationskoeffizienten zwischen $|1-\beta_i|$ und $\sum_j d_{ij}^2$ negativ und zwischen β_i und $\sum_j d_{ij}^2$ positiv, wobei die numerischen Werte jedoch durchweg sehr klein sind. Sie erweisen sich in keinem Fall als statistisch signifikant von Null verschieden. Für die üblichen Korrelationskoeffizienten, deren absolute numerische Werte in sämtlichen Fällen kleiner als die der Rangkorrelationskoeffizienten sind, erhält man völlig analoge Ergebnisse. Die für das vorliegende Datenmaterial nachgewiesene Unkorreliertheit zwischen β_i und $\sum_j d_{ij}^2$ bzw. $|1-\beta_i|$ und $\sum_j d_{ij}^2$ steht also in Übereinstimmung mit den aus der Literatur bekannten Ergebnissen einer unabhängigen Vererbung dieser beiden Parameter.

5 Diskussion

Bei der Analyse von Genotyp-Umwelt-Wechselwirkungen im Rahmen der Entwicklung quantitativer Maßzahlen zur Erfassung der phänotypischen Stabilität wird beim üblichen Regressionsansatz die Leistung x_{ij} des Genotyps i in der Umwelt j in Beziehung gesetzt zur Leistungsfähigkeit u_i der Umwelt j : $x_{ij} = \beta_i u_j + c_i + d_{ij}$. Bei der Parameterschätzung über die Bedingung » $\sum_j d_{ij}^2 = \text{Minimum}$ « werden sämtliche Abweichungen d_{ij} völ-

lig gleichwertig behandelt – trotz zahlreicher bekannter Ursachen für starke Unterschiedlichkeiten. Eine gewisse Berücksichtigung dieser Effekte durch Einführung verschiedener Gewichte w_{ij} führt zur Parameterschätzung über die Bedingung » $\sum_j d_{ij}^2 w_{ij} = \text{Minimum}$ «. Dieses Vorgehen scheint nicht nur wohlbegründet, sondern auch erforderlich zu sein. Trotzdem resultierte die Heranziehung einer ganzen Reihe von verschiedenen Gewichtungsfaktoren w_{ij} bei einem Datenmaterial einer Versuchsserie des Bundessortenamtes in mehr oder weniger identischen Rangfolgen der Sorten bezüglich Stabilität, d.h. die Verwendung einer gewogenen Regression war – verglichen mit dem üblichen ungewichteten Ansatz – praktisch ohne jeden Einfluß. Dieses könnte entweder auf den speziellen verwendeten Datensatz oder aber auf die hier gegebene Definition der Gewichtungsfaktoren zurückzuführen sein.

Diesem ersten Argument (Datensatz) kommt jedoch sicher keine besondere Bedeutung zu, denn das verwendete numerische Beispiel wurde relativ zufällig aus einem großen Datenmaterial des Bundessortenamtes ausgewählt und dürfte die realistischen Verhältnisse bei praktischen Anwendungen mehr oder weniger gut repräsentieren. Bei jeder einzelnen der hier verwendeten Definitionen der Gewichtungsfaktoren würde ein Einfluß unterschiedlicher Gewichte w_{ij} auf die Stabilitätsrangfolge der Sorten nur bei extrem unterschiedlichen numerischen Werten für die Gewichte w_{ij} zustandekommen. Solche extrem unterschiedlichen numerischen Werte sind jedoch sehr unrealistisch. Denn alle w_{ij} werden hier ja über die Ausgangsdaten x_{ij} berechnet und da das Anwendungsbeispiel die realistischen Verhältnisse bei praktischen Anwendungen repräsentieren dürfte, sind auch die hier angewendeten numerischen Werte der Gewichte zumindest größenordnungsmäßig doch als mehr oder weniger repräsentativ anzusehen – und sie führten in sämtlichen untersuchten Fällen ja zu praktisch identischen Sortenrangfolgen bezüglich Stabilität verglichen mit der Rangfolge im Fall »ungewichtet«. Zu untersuchen wäre daher als nächstes die Frage, ob und in welcher Weise die Gewichtungsfaktoren eventuell anders definiert werden müßten, um zu einem stärkeren Einfluß und damit zu einer deutlicheren Differenzierung bei der Stabilitätsrangfolge der Sorten zu führen. Dieses soll in späteren Untersuchungen geschehen.

Eine Hauptschwierigkeit bei der Anwendung des Regressionsansatzes zur Charakterisierung der phänotypischen Stabilität besteht in einer sinnvollen Festlegung und dann Schätzung der Bezugsbasis, d.h. der Leistungsfähigkeit u_j der Umwelt j . Die verschiedenen verwendeten Bezugsbasen führten zum Teil zu extrem unterschiedlichen Sortenrangfolgen bezüglich Stabilität, wobei jedoch das zugrundegelegte Stabilitätskonzept (maximale Stabilität bei $\beta = 0$ oder $\beta = 1$) eine besondere Rolle spielt (siehe Kap. 4). Lediglich arithmetischer Mittelwert, Dichtemittel, Zentralwert sowie arithmetisches Mittel der Standardgruppe zeigen jeweils untereinander keine allzugroßen Unterschiede in den resultierenden Sortenrangfolgen bezüglich Stabilität. Dies ist verständlich, da die genannten Mittelwertsmaße »normalerweise« mehr oder weniger dicht beieinander liegen. Nur bei stark asymmetrischen Häufigkeitsverteilungen können größere Unterschiede auftreten. Wenn das Sortiment genügend groß ist und wenn die verwendeten Genotypen als mehr oder weniger repräsentative Stichprobe des gesamten Genotypenspektrums angesehen werden können, trifft diese Voraussetzung jedoch nicht zu.

Das anfangs kurz diskutierte Stabilitätskonzept, wonach die Leistungsstabilität unabhängig von der Leistungshöhe sein sollte, wurde bei verschiedenen der beschriebenen Ansätze nicht vollständig eingehalten. Denn die Stabilitätsparameter

hängen von den Gewichten ab und diese wiederum wurden ja in verschiedenen Fällen in einem deutlichen Bezug zur Leistungshöhe definiert – wie z. B. bei $w_{ij} = x_{ij}$. Damit findet eine Vermengung von Merkmalshöhe- und Merkmalsstabilitätsgesichtspunkten bei der Stabilitätsfeststellung statt.

Bevor irgendwelche endgültigen Schlußfolgerungen und Festlegungen sowohl hinsichtlich der Auswahl von bestimmten Bezugsbasen als auch bezüglich der Verwendung der verschiedenen Gewichtsfaktoren getroffen werden können, sollten zunächst erst einmal die weiteren Erfahrungen bei der Anwendung solcher Regressionsrechnungen auf ein möglichst umfangreiches und möglichst verschiedenartiges Datenmaterial abgewartet werden. Diese empirischen Untersuchungen werden sicher zusätzliche Informationen liefern und Entscheidungen über die anzuwendenden optimalen Bezugsbasen und Gewichte ermöglichen. Solche Anwendungen auf ein äußerst umfangreiches Datenmaterial für verschiedene landwirtschaftliche Kulturpflanzen (Winterweizen, Mais, Raps, Zuckerrüben) sind in Vorbereitung.

Literaturverzeichnis

- BUSCH, R. H., HAMMOND, J. und R. C. FROHBURG (1976): Stability and performance of hard red spring wheat bulks for grain yield. *Crop Science* 16, 256–259.
- DHILLON, B. S. und J. SINGH (1977): Estimation and inheritance of stability parameters of grain yield in maize. *J. Agric. Sci., Camb.* 88, 257–265.
- EBERHART, S. A. und W. A. RUSSELL (1969): Yield and stability for a 10-line diallel of single-cross and double-cross maize hybrids. *Crop Science* 9, 357–361.
- FREEMAN, G. H. und J. M. PERKINS (1971): Environmental and genotype-environment components of variability. VI. II. Relations between genotypes grown in different environments and measures of these environments. *Heredity* 27, 15–23.
- FRIPP, Y. J. (1972): Genotype-environment interactions in *Schizophyllum commune*. II. Assessing the environment. *Heredity* 28, 223–238.
- FREEMAN, G. H. (1973): Statistical methods for the analysis of genotype-environment interactions. *Heredity* 31, 339–354.
- FATUNLA, T. und K. J. FREY (1976): Repeatability of regression stability indexes for grain yield of oats. *Euphytica* 25, 21–28.
- FISCHBECK, G. (1976): Beiträge zur quantitativen Differenzierung der Ertragsicherheit. Vortrag 20. Jahrestag. Ges. Pflanzenbauwiss. Weihenstephan 11. 10. 1976 (Kurzfassungen der Vorträge 4–5).
- GEIDEL, H. (1971): Zur Interpretation von Interaktionen bei Versuchsserien. *EDV in Medizin und Biologie* 2, 13–18.
- HANSON, W. D. (1970): Genotypic stability. *Theoretical and Applied Genetics (TAG)* 40, 226–231.
- HARDWICK, R. C. und J. T. WOOD (1972): Regression methods for studying genotype-environment interactions. *Heredity* 28, 209–222.
- HAUFE, W. und H. GEIDEL (1978a): Zur Beurteilung der Ertragsicherheit von Sorten und Zuchtstämmen. I. Definitionen, Stabilitätsparameter und deren Interpretationsmöglichkeiten. *Zeitschr. f. Pflanzenzüchtg.* 80, 24–37.
- HAUFE, W. und H. GEIDEL (1978b): Zur Beurteilung der Ertragsicherheit von Sorten und Zuchtstämmen. II. Der Einsatz von Stabilitätsparametern in der Praxis. *Zeitschr. f. Pflanzenzüchtg.* 81, 1–16.
- HAUFE, W. und H. GEIDEL (1978c): Probleme bei der Zusammenfassung von Einzelversuchen zu Versuchsserien. *EDV in Medizin und Biologie* 9, 12–20.
- JENSEN, N. F. (1976): Floating checks for plant breeding nurseries. *Cereal Res. Comm.* 4 (3), 285–295.
- LEIN, A. und W. HAUFE (1967): Zur Frage der Bezugsbasis für statistische Maßzahlen zur Beurteilung der ökologischen Streubreite. *AG Biometrie in der Pflanzenzüchtung (DLG); Rundschreiben* 2/1967, 2–12.
- MÜNZER, W. (1972): Untersuchungen zur Charakterisierung der ökologisch bedingten Variabilität verschiedener Zuchtsorten und der Differenzierungseigenschaften verschiedener Prüforte. *Zeitschr. f. Pflanzenzüchtg.* 67, 337–352.
- PERKINS, J. M. und J. L. JINKS (1968): Environmental and genotype-environment components of variability. IV. Non-linear interactions for multiple inbred lines. *Heredity* 23, 525–535.
- RUNDFELDT, H. (1980): Biometrische Überlegungen zur Errechnung optimaler Generalmittel. *Proceedings Interregionales Biometr. Koll. München* 1980, 54–60.
- SCHNELL, F. W. (1967): Die Methoden zur Erfassung der phänotypischen Stabilität von Zuchtsorten. *AG Biometrie in der Pflanzenzüchtg. (DLG), Rundschreiben* 1/1967, 18–21.
- SEIF, E. und D. G. PETERSON (1978): Effect of rainfall on the grain yield of spring wheat with an application to the analysis of adaptation. *Austr. J. Agric. Res.* 29, 1107–1115.
- UTZ, H. F. (1972): Die Zerlegung der Genotyp-Umwelt-Interaktionen. *EDV in Medizin und Biologie* 3, 52–59.
- WRICKE, G. (1965): Die Erfassung der Wechselwirkung zwischen Genotyp und Umwelt bei quantitativen Eigenschaften. *Zeitschr. f. Pflanzenzüchtung* 53, 266–343.
- WELING, F. (1973): Zur Bestimmung der Leistungsstabilität von Pflanzentypen nach den Verfahren von WRICKE, FINLAY, EBERHART u. a. Zugleich ein Beitrag zum Wesen sowie zur Analyse von Wechselwirkungen. *EDV in Medizin und Biologie* 4, 88–98.
- WOLFRAM, H. (1973): Methoden zur Untersuchung von Genotyp-Umwelt-Wechselwirkungen in der Pflanzenzüchtung. *Arch. Züchtungsforsch.* 3(1), 85–106.
- WRICKE, G. und W. E. WEBER (1980): Erweiterte Analyse von Wechselwirkungen in Versuchsserien. *Proceedings Interregionales Biometr. Koll. München* 1980, 87–95.

Eingegangen am 21. 7. 80

Anschrift des Verfassers: Prof. Dr. M. Hühn, Universität Kiel, Institut für Pflanzenbau und Pflanzenzüchtung, Olshausenstraße 40–60, D-2300 Kiel.

EDV-Programm zu stochastischen Modellen der Karzinogenese am Beispiel der Hepatozytenentwicklung

W. Hopfenmüller

Zusammenfassung

In dieser Arbeit geht es darum, die Bestandteile des Zellplasmas der Zygote auf ihrer »somatischen Entwicklungsbahn« bis hin zum Organ zu beobachten. Für die Mitochondrien – die für die extrachromosomale Vererbung entscheidenden Zellorganellen – wurde am Beispiel der Leberzellentwicklung ein Organentwicklungsmodell mit seinen Computerprogrammen entwickelt.

Summary

The cytoplasm of a zygote is observed on the somatic way of evolution up to the organ. For mitochondria, which are the most important cell components for the extranuclear heredity, a model of evolution for liver cells with its computer programs was developed.

Basierend auf experimentellen Ergebnissen wurde vor einigen Jahren postuliert, daß mitochondriale DNA und der mitochondriale Stoffwechsel die primären Zielscheiben für einige Zytostatika und demnach auch die primären Ursachen für den Prozeß der Karzinogenese sind (Neubert et al., 1975). Dafür konnte eine Hypothese ausgearbeitet werden, die eine überzeugendere Erklärung der Karzinogenese liefert als Arbeiten, die auf der Grundlage einer Anreicherung von Veränderungen des nuklearen Erbgutes beruhen (Neubert, Hopfenmüller, Fuchs, 1980). Die mathematischen Modelle für solch eine Möglichkeit einer extrachromosomalen Vererbung wurden bereits vorgestellt und dabei auf eine Weiterführung mit Hilfe der Mitochondrientheorie auf somatische Zellen hingewiesen (Hopfenmüller, 1978, 1979 b). Die dazugehörigen EDV-Programme beschränkten sich bisher auf die weibliche Keimbahn, da Teilungsgewohnheiten und Lebensdauer von somatischen Zellen sehr vielfältig sind und genaue Zahlen oft nicht angegeben werden können (Hopfenmüller, 1979 a).

In Übereinstimmung mit den ursprünglichen, antreibenden Mechanismen auf der weiblichen Keimbahn wird die Karzinogenese, hier speziell in den Leberzellen, verstanden als Zweistufenprozeß:

- 1) die Ursprungssituation veränderter mitochondrialer DNA und
- 2) ein rein stochastischer Prozeß, führend zu einer Auswahl von Zellen, in der die Läsion manifest wurde.

Für diese somatischen Modelle wurden folgende Annahmen gemacht (Neubert, Hopfenmüller, Fuchs, 1980):

- 1) Die DNA-Moleküle einiger Mitochondrien haben sich geändert, wodurch sogenannte »pathologische« Zellorganellen neben den normalen Mitochondrien einer Leberzelle entstehen.
- 2) Während der Zellteilungsperiode werden die pathologischen Mitochondrien zufällig auf die Tochterzellen verteilt.
- 3) Ist ein bestimmter Grad der Verschiebung von normalen zu »pathologischen« Mitochondrien erreicht, so ändert die Zelle ihre Charakteristik und »benimmt sich« wie eine Tumorzelle.

Das hier vorgestellte Modell ist für Hepatozyten zugeschnitten. Für andere Typen von Leberzellen, wie Endothelialzellen, Gallengangszellen oder »von Kupfersche Sternzellen«, weichen die Bedingungen von den Modellvoraussetzungen ab. Es wird angenommen, daß die menschliche Leberzelle ca. 1600 Mitochondrien enthält. Die Dauer einer sogenannten »mitotischen« Generation läßt sich schwer exakt bestimmen, doch geht man heute davon aus, daß eine Generation ca. 200 Tage dauert. So kann man etwa mit einem Maximum von 100 Generationen in einem Lebensalter rechnen. Die 30000 Mitochondrien der menschlichen Eizelle werden bei den Furchungen der Präimplantationsperiode mit gleichen Anteilen auf die Tochterzellen verteilt, so daß sich die Zahl der Mitochondrien in einer Blastozyste auf etwa 300 reduziert hat. Bis etwa zum 18. Lebensjahr ist die Leberzelle dem Zellwachstum unterworfen gekoppelt mit normalen mitotischen Teilungen. Die bei der Eizelle wesentliche Wachstumsperiode fällt also hier in die erste Phase der mitotischen Teilungen.

Für das Modell wurde die Annahme gemacht, daß nach einer erfolgten mitotischen Zellteilung die Tochterzellen um 50 Mitochondrien anwachsen, bevor die nächste Zellteilung einsetzt. Auf diese Weise kommt man auf 27 »mitotische« Generationen (200 Tage pro Generation) in den ersten 18 Lebensjahren (Anwachsen von 300 Mitochondrien auf ca. 1600 Mitochondrien).

Für die i-te Hepatozytenentwicklung gelten folgende Ansätze (Hopfenmüller, 1979 b)

a) für die i-te Teilung + i-te Verdoppelung

$$p_{i,(i-1)v}(m|n+(i-1)v) = \sum_{\beta=0}^{n+(i-1)v} p_{i-1,(i-1)v}(\beta|n+(i-1)v) \frac{\binom{2\beta}{m} \binom{2(n+(i-1)v)-2\beta}{n+(i-1)v-m}}{\binom{2(n+(i-1)v)}{n+(-1)v}}$$

b) für das i-te Wachstum

$$p_{i,iv}(m|n+iv) = \sum_{\alpha=m-1}^m p_{i,iv-1}(\alpha|n+iv-1) \left(1 - \frac{\alpha}{n+iv-1}\right)^{|1-\alpha-m|} \left(\frac{\alpha}{n+iv-1}\right)^{|\alpha-m|}$$

Zur Erläuterung und zur besseren Übersicht wird die Wahrscheinlichkeitsmatrix wie folgt angegeben (sie wird bei der Rechnung zeilenweise abgearbeitet):

$$\begin{pmatrix} p_{1,0}(m|n) & p_{1,1}(m|n+1) \\ p_{2,v}(m|n+v) & p_{2,v+1}(m|n+v+1) \\ \vdots & \vdots \\ p_{i,(i-1)v}(m|n+(i-1)v) & p_{i,(i-1)v+1}(m|n+(i-1)v+1) \\ \vdots & \vdots \\ \dots p_{1,v}(m|n+v) \\ \dots p_{2,2v}(m|n+2v) \\ \vdots \\ \dots p_{i,iv}(m|n+i) \end{pmatrix}$$

Neu ist bei der Leberzelle die Leberzellenentwicklungsperiode (Programm LZEW). Bei der vor dieser Entwicklungsphase stattfindenden Zellfurchung und der nach dem 18. Lebensjahr spontan einsetzenden Mitosen, sowie den TURN-OVER-Prozessen sind die Programme FURCH, VERMEHR und TURN bereits vorgestellt worden (Hopfenmüller 1979 a).

```
PROGRAM LZEW
DATA IN/97/, LOUT/99/
DIMENSION V(3010), W(3010), D(3010)
```

```
C AN=N, A: STARTWERT, L: MAXIMALE GENERATIONSAN-
C ZAHL
C J: UEBERLAUFZAEHLER, S=P(M/N), HYP: KOEFFIZIENT
C DER
C MODIFIZIERTEN HYPERGEOMETRISCHEN VERTEILUNG
```

```
C AN=N, A: STARTWERT, LL: ANZAHL DER MITOCHON-
C DRIENTEILUNGEN
C ZET: AUSGABEBESCHRAENKUNG, MZ, IS: RECHENPARA-
C METER
```

```
L=0
A=10
AN=300
K=AN
KK=K+1
IA=A
```

```
1 L=L+1
K=AN
IK=AN/2
S=0.
HYP=1.
J=0
LL=50
ZET=.100000E-13
```

```
MZ=0
IS=0
KG=1
SU=0.
```

```
C
C UEBER- UND UNTERLAUFANZEIGE
C
C CALLOVERFL(J)
```

```
C
C EIN- UND AUSGABEFORMATE
```

```
10 FORMAT(7X,I4,7X,E13.6,9X,F10.7)
20 FORMAT(1H,7X,'-----',/)
30 FORMAT(7X,'N='F6.0,3X,'A='F4.0,2X,'BEI',I3,2X,1'GENE-
Rationen UND',I6,2X,'MITOCHONDRIENTEILUNGEN'//)
40 FORMAT(10X,I3,'.GENERATION'/)
50 FORMAT(10X,'MITOCHONDRIENTEILUNGEN:
1',I6,3X,'N='F6.0,3X,'A='F4.0,2X,'BEI',I3,2X,1'GENE-
Rationen UND',I6,2X,'MITOCHONDRIENTEILUNGEN'//)
60 FORMAT(E13.6)
70 FORMAT(I6)
80 FORMAT(F6.0)
```

```
C
C ALLGEMEINE UEBERSCHRIFT
C
C WRITE(LOUT,30)AN,A,L,LL
```

```
C
C EIN- UND AUSGABE DER 0-TEN GENERATION
```

```
55 DO 4715 IP=1, KK
IF(IP-1-IA)3,4,3
3 D(IP)=0.
GO TO 31
4 D(IP)=1.
31 IL=IP-1
WRITE(LOUT,10)IL,D(IP),D(IP)
4715 CONTINUE
```

```
C
C AUSGABE VON - NACH BEENDIGUNG EINER GENERA-
C TION
C
C WRITE(LOUT,20)
```

```
C
C BERECHNUNGEN DER WAHRSCHEINLICHKEITEN FUER
C LGENERATIONEN
```

```
C
C UEBERSCHRIFT
```

```
11 WRITE(LOUT,40)L
```

```
C
C UMSPEICHERSCHLEIFE
```

```
66 DO 4716 II=1, KK
W(II)=D(II)
4716 CONTINUE
```

```
C
C I: ZAEHLER FUER ERRECHNETE PATHOLOGISCHE MITO-
C CHONDRIEN
C BERECHNUNG DER SUMMATIONSGRENZEN
```

```
22 DO 4712 IJ=1, KK
KU=(IJ+2)/2
KRU=(IJ+1)/2
KRO=IK+KRU
KO=MIN0(KRO, KK)
```

```
C
C BERECHNUNG DER KOEFFIZIENTEN
```

```
33 DO 4713 I=KU, KO
IF(W(I))23,23,38
38 XBM=1+I-2
X2N=K+K
X2B=X2N-XBM
ML=IJ-1
NMM=K-(IJ-1)
KN=K
IF(IJ-1)707,707,709
```

```
C
C VERZWEIGUNGSSCHLEIFE FUER IJ=1
```

```
707 DO 4714 IM=1, K
F=X2B/X2N
HYP=HYP*F
X2B=X2B-1
X2N=X2N-1
```

```
4714 CONTINUE
C
C BERECHNUNG DER SUMMANDEN
```

```
B=W(I)*HYP
GO TO 99
```



```

C 709 IF(IJ-KK)505,808,808
C
C 505 DO 4724 IM = 1,K
X1F = XBM*KN
X2F = ML*X2N
F = X1F/X2F
HYP = HYP*F
XBM = XBM-1
KN = KN-1
ML = ML-1
X2N = X2N-1
C
C VERZWEIGUNG WEGEN BUEBERØ
C
C IF(ML)67,67,4724
67 XBM = X2B
ML = NMM
4724 CONTINUE
C
C BERECHNUNG DER SUMMANDEN
C
C B = W(I)*HYP
GO TO 99
C
C BERECHNUNG FUER IJ = N + 1
C
808 X2B = XBM
GO TO 707
C
23 B = Ø.
C
C BERECHNUNG DER SUMME: P(M/N)
C
99 S = S + B
HYP = 1.
C
4713 CONTINUE
C
C AUSGABE DES ERRECHNETEN WERTES
C
V(IJ) = S
LJ = IJ-1
IF(V(IJ))71,71,72
72 WRITE(LOUT,10)LJ,V(IJ),V(IJ)
71 S = Ø.
C
4712 CONTINUE
C
C AUSGABE VON --- NACH BEENDIGUNG EINER GENERA-
TION
C
WRITE(LOUT,20)
C
LV = Ø
C
5 DO 4717 IB = 1, KK
IF(V(IB))501,501,502
502 LV = LV + 1
4717 CONTINUE
C
501 KV = LV-1
C
2 DO 4711 IC = 2, LV
V(IC-1) = V(IC)
4711 CONTINUE
C
12 DO 5711 IUM = 1, LL
KM = IUM + K
IF(KG-2)701,701,702
702 IF(KG-48)703,703,701
701 WRITE(LOUT,50)IUM,KM
703 IV = KV + IUM
N = K + IUM-1
FN = FLOAT(N)
C
C BERECHNUNG DER P(M/N + K)
C
IT: ZAEHLER FUER ERRECHNETE PATHOLOGISCHE MITO-
CHONDRIEN
C
21 DO 5712 IT = 1, IV
M = IT
IF(M-IV)17,34,34
17 FI = FLOAT(M)
B1 = V(M)*(1.-FI/FN)
GO TO 37
34 B1 = Ø.
37 M = M-1
IF(M-1)36,18,18
18 FI = FLOAT(M)
B2 = V(M)*(FI/FN)
GO TO 47
36 B2 = Ø.
47 C = B1 + B2

```

```

C BESCHRAENKUNG DER AUSGABE
C
IF(KG-50)108,107,107
107 IF(C-ZET)49,48,48
108 IF(C-ZET)49,51,51
48 WRITE(LOUT,10)IT,C,C
IS = IS + 1
GO TO 51
C
C UMSCHAUFELUNG DER WERTE
C
49 MZ = MZ + 1
51 U = C
IF(IT-1)13,13,15
15 JJ = IT-1
V(JJ) = Q
13 Q = U
IF(MZ-3)5712,5712,909
C
5712 CONTINUE
909 KW = JJ + 1
V(KW) = Q
MZ = Ø
C
IF(IS)6,6,8
6 KG = KG + 1
GO TO 7
8 KG = 1
IS = Ø
C
C AUSGABE VON --- NACH BEENDIGUNG EINER GENERA-
TION
C
C (BESCHRAENKUNG DER AUSZUDRUCKENDEN WERTE)
C
7 IF(KG-3)607,607,708
708 IF(KG-48)5711,5711,607
607 WRITE(LOUT,20)
C
5711 CONTINUE
C
14 DO 5713 IH = 1, KW
SU = SU + V(IH)
5713 CONTINUE
C
VSU = 1.-SU
C
16 DO 5714 JH = 1, KW
D(JH + 1) = V(JH)
5714 CONTINUE
C
D(1) = VSU
KK = KW + 1
AN = AN + LL
IF(K-1600)1,9999,9999
PROGRAMMENDE
9999 STOP
END

```

Literaturverzeichnis

- BIELKA, H.: 1973: Molekulare Biologie der Zelle. Gustav-Fischer Verlag, Stuttgart
- HOPFENMÜLLER, W.: 1978: A mathematical model for extrachromosomal heredity, Biomed. J. 20, 609 – 618
- HOPFENMÜLLER, W.: 1979 a: Programs for the mathematical model of the problem of extrachromosomal heredity, EDV in Medizin und Biologie 10, 119 – 122
- HOPFENMÜLLER, W.: 1979 b: A mathematical model for extrachromosomal heredity in somatic cells, exemplified by liver cells, Biomed. J. 21, 431 – 437
- LINKS, J.L.: 1967: Extrachromosomale Vererbung, Gustav-Fischer Verlag, Stuttgart
- NEUBERT, D.: 1975: Occurrence and possible functions of mitochondrial DNA in animal development, Biochemistry of Animal Development, 3 Hrsg.: R. Weber, Academic Press, Inc., New York
- NEUBERT, D.: HOPFENMÜLLER, W.: FUCHS, G.: 1980: Manifestation of carcinogenesis as a stochastic process on the basis of an altered mitochondrial genome. (im Druck)
- RÖBBELEN, G.: 1966: Zytogenetik des Plasmas. Deutsche Medizinische Wochenschrift 91, 715–720.

Eingegangen am 29.9.80

Anschrift des Verfassers: Dr. rer. nat. Werner Hopfenmüller, Hochschulassistent am Institut für Medizinische Statistik und Dokumentation der Freien Universität Berlin, Hindenburgdamm 30, 1000 Berlin 45

NACHRICHTEN UND BERICHTE

BUCHBESPRECHUNGEN

SIEMENS »data praxis« Grundlagen der Softwaretechnologie

Die Entwicklung und Herstellung von Softwareprodukten hat in den vergangenen Jahren industrielle Größenordnungen angenommen. Es ist daher verständlich, daß man sich bemüht, ingenieurmäßiges Arbeiten auch unter Ausnutzung des Computers als Hilfsmittel auf die Softwareentwicklung zu übertragen.

In dem vorliegenden Heft der »data praxis« wird ein weitgehend geschlossenes Technologiekonzept für die Softwareproduktion vorgestellt.

Interessenten können diese Schrift (Best.-Nr. D 90/112-01) kostenlos bei der Fa. SIEMENS AG, Bereich Daten- und Informationssysteme, Postfach 83 29 40, 8000 München 83, anfordern.

Statistics in Medicine

8.-11. July 1981
King's College, Cambridge

Sessions have been planned on:

Evaluation of drugs, to cover aspects relating to ethics and the law, to the monitoring of adverse effects of drugs to methodological considerations of a clinical trial as well as to certain aspects of techniques of analysis.

Survival analysis, as it relates to cancer experiments, trials of myocardial infarction etc.

Computers, Statistics and Medicine, to cover the use of health statistics and the use made of computers and statistics in various screening services, in diagnosis and prognosis.

Research, to present some recent advances in the application of statistics to medical problems.

Additionally, the organising committee are looking for other papers that will be shorter than those in the main sessions and which may be featured in parallel sessions. Those willing to submit a synopsis for consideration should write to the Conference

Technical Convenor: Dr. J. Seldrup F.I.S., Ciba-Geigy Pharmaceuticals Division, Wimblehurst Road, Horsham, West Sussex RH12 4 AB.

Conference Office: Mrs. T. J. Konrath, The Institute of Statisticians, 36 Churchgate Street, Bury St. Edmunds, Suffolk IP33 1 RD

Medizin – Technik 81

15. Jahrestagung der Deutschen Gesellschaft für Biomedizinische Technik sowie Jahrestagung der Österreichischen und Schweizerischen Gesellschaft für Biomedizinische Technik.

Termin: 9.-11. September 1981

Ort: Universität Stuttgart

Die Anmeldung von Vorträgen mit einer Vortragskurzfassung muß bis zum 30.04.1981 erfolgen.

Weitere Informationen und Formblätter für die Vortragskurzfassung erhalten Sie durch die Tagungsgeschäftsstelle MEDIZIN-TECHNIK 81, Universität Stuttgart, Postfach 560, D-7000 Stuttgart 1

SNEED, H. M.
Software-Entwicklungsmethodik
1980, 196 S., DM 42,-
Verlagsges. R. Müller, Köln

Es ist unübersehbar, daß die Bedeutung der Software immer mehr zunimmt. Daher ist nur zu verständlich, daß man sich Gedanken über eine ingenieurmäßige Software-Entwicklung macht.

Das vorliegende Buch ist eine ausgezeichnete Darstellung, die versucht, in dieses zur Zeit noch unübersichtliche Gebiet Systematik und Klarheit zu bringen. Dieser Versuch muß als gelungen angesehen werden. – Die Aufzählung der Kapitelüberschriften: Software-Technologie, Software-Systemspezifikation, Software-Entwurf, Programmier-technik, Software-Qualitätssicherung, Software-Systemintegration und Modell eines Software-Entwicklungssystems, vermag nur zum Teil den wirklichen Wert dieses Buches zu beschreiben. Ein ausführliches Literaturverzeichnis ergänzt die klar gegliederten und informativen Ausführungen. Ein Buch, das jeder Programmierer verfügbar haben sollte. Ge.

KATZAN, H.
Methodischer Systementwurf
Eine Einführung in die HIPO-Technik
übersetzt und bearbeitet von M. Bues
1980, 201 S., DM 42,-
Verlagsges. R. Müller, Köln

Die Beschreibung von Systemen ist heute nicht nur ein Teil der Informationsverarbeitung. Hier ist sie aber unbedingt notwendig. Daher ist es nicht verwunderlich, daß immer wieder neue Vorschläge auftauchen, mit denen Systeme beschrieben werden können und die möglichst zwangsläufig auch zu einer brauchbaren Dokumentation führen. Das in diesem Buch beschriebene HIPO-Verfahren (Hierarchie plus Input-Prozess-Output) ermöglicht eine graphische Darstellung von Funktionen, die von einem System ausgeführt werden und stellt die Beziehungen zwischen den Eingaben, den Verarbeitungsstufen und Ausgaben innerhalb eines Systems dar. – Nach einer allgemeinen Einführung wird die HIPO-Technik behandelt und an einer Fallstudie erläutert. Danach werden noch verschiedene Darstellungstechniken, Standards sowie der computergestützte Einsatz angesprochen. Ge.

BARRITT, M. M. and WISHART, D. (Ed.)
COMPSTAT 1980
Proceedings in Computational Statistics
4th Symposium held at Edinburgh 1980
1980, 632 S., DM 68,-
Physica-Verlag, Würzburg

Das vierte COMPSTAT Symposium zeigte wieder, daß die Probleme im Zusammenspiel von statistischen Methoden und der Benutzung von Computern recht vielfältig sind. So ergeben die in den Proceedings zusammengestellten Referate einen guten Überblick über vorhandene Softwarepakete und auftretende Probleme beim Umsetzen von statistischen Methoden in Programme. An Hauptthemen wurden u.a. behandelt: Stichprobenverfahren, Datenbankprobleme, Varianz-, Kovarianzanalysen, interaktives Rechnen, lineare und nichtlineare Regression, multivariate Analysen, Optimierungen und Simulationen, Clusteranalysen, statistische Software und Zeitreihenanalysen. Daneben beschäftigten sich Referate mit der Datenanalyse bei großen Datenmengen und mit einer Taxonomie der statistischen Software. Die Herausgeber sind zu beglückwünschen, daß dieser instruktive Band, wie auch bei den vorhergehenden Veranstaltungen, zum Kongreß vorlag. Ge.

KRÖTLINGER, M.
Einführung in die elektronische Datenverarbeitung in der Allgemeinpraxis
1980, 90 S., DM 28,-
Verlag W. Maudrich, Wien – München – Bern

Die vorliegende Broschüre ist aus der praktischen Arbeit des Autors hervorgegangen und stellt eine nützliche Diskussionsgrundlage für die hier entstehenden Probleme dar. Bei den fallenden Hardwarepreisen

sollte man sich intensiver mit der gedanklichen Durchdringung dieses Fragenkomplexes beschäftigen. Für diese zweifellos notwendigen Bemühungen kann die vorliegende Einführung sehr wertvoll sein. – Nach einer Darstellung der zu verarbeitenden Daten werden kurz einige Grundlagen der Computer behandelt. Nach einer Ist-Analyse der Allgemein-Praxis, folgt die Beschreibung eines Programmes. Ein Ausblick auf mögliche Datennetze sowie Anwendungsbeispiele von Kleinrechnern in der Allgemeinpraxis ergänzen die Darstellung. Ge.

SCHÄFER, P. und WICZORKE, M.
Lexikon der Prozeßrechner-technik
1979, 327 S.
Siemens AG, Berlin und München

Das vorliegende Lexikon enthält in alphabetischer Reihenfolge sowohl Begriffe mit Erläuterungen aus der Hardware und der Software der Prozeßrechner-technik als auch der Meß-, Steuerungs- und Regel-technik und der Prozeßautomatisierung. Dabei wurden die zur Zeit gültigen DIN-Normen berücksichtigt. Dem jeweiligen Begriff folgt unmittelbar seine englische Übersetzung. Im Anhang sind die englischen Begriffe mit den entsprechenden deutschen Übersetzungen alphabetisch aufgeführt. Da auch die Begriffe der normalen Datenverarbeitung erfaßt sind, liegt hier ein wirklich nützliches Nachschlagewerk vor. Ge.

GULBINS, J.
Einführung in die Benutzung der PDP 11 mit Betriebssystem RSX 11 M
Band 1:
1979, 192 S., DM 16,80
Band 2:
1979, 202 S., DM 16,80
Neuer Verlag B. Bruscha, Tübingen

Wenn sich diese Darstellung auch nur auf die PDP 11 bezieht, so könnte sie doch ein Beispiel sein auch für andere Rechner. Die Fülle von Informationen des Herstellers wurde aufgearbeitet und verdichtet zu einem brauchbaren Handbuch. Für PDP 11 Benutzer ein nützliches Handbuch, das aber auch anderen Computerbenutzern gute Einblicke in den Aufbau und die Arbeitsweise eines Computers vermittelt. Ge.

LE ROY, H. L.:
Prinzipien der linearen Algebra
UTB Bd. 984
1980, 328 S., DM 26,80
Verlag Paul Haupt, Bern

Die Verquickung von statistischen Auswertungsmethoden und linearer Algebra läßt sich nicht mehr übersehen. Dabei bedarf es für das Verständnis vieler statistischer Methoden solider Kenntnisse der Prinzipien der linearen Algebra.

Es ist das Verdienst des Autors, daß er genau aus dieser Blickrichtung die notwendigen Hilfsmittel bereitgestellt und an überschaubaren Beispielen erläutert hat. Hervorzuheben sind die systematische Darstellung der jeweils möglichen Fälle sowie der Versuch, soweit wie möglich, graphische Darstellungen für die Erläuterungen der Sachverhalte zu verwenden.

Das Buch wird vielen helfen, sich in die Prinzipien der linearen Algebra einzuarbeiten und damit die Möglichkeiten und Einzelheiten statistischer Verfahren besser zu verstehen. Ge.

SCHAEFER, J.
Tischcomputer – BASIC für Anfänger
1979, 199 S., DM 24,80
R. Oldenbourg Verlag, Wien-München

Der zunehmenden Verbreitung der mit BASIC zu programmierenden Tischcomputer kommt die vorliegende Einführung entgegen. Daß die Mehrzahl der Beispiele aus dem Bereich der Schulmathematik bzw. Technik gewählt wurde, sollte einer Benutzung in anderen Bereichen nicht entgegenstehen. Die Darstellung ist ausgerichtet auf den Tischcomputer TRS 80. Die eigentliche Einführung in BASIC gilt aber weitgehend auch für ähnliche Tischcomputer. Ge.

Biomathematics Vol. 10

OKUBO, A.

Diffusion and Ecological Problems: Mathematical Models

1980, 254 S., DM 78,-

Springer-Verlag, Berlin – Heidelberg – New York

Das vorliegende Buch ist eine erweiterte Version der japanischen Ausgabe »Ecology and Diffusion« von R. SHOKAN (1975). Die Bedeutung dieses Fachgebietes zeigt sich u.a. in der Vielzahl der stofflichen Ergänzungen seit 1975, die sich auch in der doppelten Anzahl von Literaturhinweisen niederschlagen.

Die behandelten Probleme geben einen guten Überblick über die Möglichkeiten der Anwendung von mathematischen Modellen in der Ökologie und zeigen gleichzeitig die Vielfalt der Fragestellungen auf. Für Interessenten ein ausgezeichnetes Lehrbuch. Ge.

STEINHAUSEN, D. und LANGER, K.

Clusteranalyse

Einführung in Methoden und Verfahren der automatischen Klassifikation

1977, 206 S., DM 34,-

W. de Gruyter, Berlin-New York

»Mit zahlreichen Algorithmen, FORTRAN-Programmen, Anwendungsbeispielen und einer Kurzdarstellung der nichtvariablen statistischen Verfahren«, so wird im Untertitel der Inhalt dieses Buches sehr treffend charakterisiert. Die Verfasser haben an mehreren Stellen den Versuch unternommen, unterschiedliche Methoden durch Verwendung geeigneter Parameter mit Hilfe einheitlicher Formeln zu beschreiben. Dadurch wird vieles überschaubarer in dieser oft recht unübersichtlichen Materie. Es muß begrüßt werden, daß die Autoren immer wieder auf die Nicht-Eindeutigkeit der möglichen Lösungen hinweisen. – Insgesamt eine sehr gute und nützliche Zusammenstellung. Ge.

BURHENNE, W.E. und PERBAND, K. (Hrg.)

EDV-Recht

Systematische Sammlung der Rechtsurkunden, organisatorischen Grundlagen und Entscheidungen zur elektronischen Datenverarbeitung

Ergänzende Ausgabe, einschl. 25. Lieferung (März 1980), 2.286 Seiten und 3 Ausschlagtafeln, DIN A 5, DM 74,- zuzüglich je DM 10,80 für 2 Spezialordner.

Erich Schmidt Verlag, Berlin-Bielefeld-München

Zu diesem umfassenden Werk für das EDV-Recht ist inzwischen die 25. Lieferung herausgekommen. Die gewählte Form (Lose-Blattsammlung) ermöglicht es, daß nur die jeweils noch aktuelle Information aufbewahrt wird. Der Wert dieser Sammlung liegt neben der Sammlung der relevanten Gesetze und Verordnungen vor allem in der Sammlung von Kommentaren zu den entsprechenden Gesetzen und Verordnungen. In der letzten Zeit sind neben den Datenschutzgesetzen auch die »Besonderen Vertragsbedingungen« mit aufgenommen worden. – Es ist erstaunlich, daß diese Sammlung so preiswert erscheinen kann. Ge.

ENDERLEIN, W.

Projektiertung von Anwenderdatennetzen

1979, 287 S., DM 87,-

R. Oldenbourg Verlag, München-Wien

Die Datenfernverarbeitung gewinnt immer mehr an Bedeutung. Daher ist es unumgänglich, die Netze sorgfältig und kostengünstig zu planen. Hier ist das vorliegende Buch, in dem die möglichen Einflußgrößen systematisch behandelt werden, eine große Hilfe. Die dargestellten Normogramme können für jeden Einzelfall zumindest erste Abschätzungen liefern. Dabei ist bewußt der Dialogdatenverkehr neben dem Batchdatenverkehr behandelt worden. Das Buch ist daher ein unentbehrliches Hilfsmittel für den Planer, aber gleichzeitig auch eine gute Einführung in diese nicht immer einfachen Sachverhalte. Ge.

RIENHOFF, O. and ABRAMS, M.E. (Ed.)

The Computer in the Doctor's Office

Proceedings of the IFIP-IMIA (TC-4)

Working Conference, Hannover, 25. – 29. April 1980

1980, 352 S., US \$ 44.00

North-Holland Publishing Comp., Amsterdam and New York

Die Bedeutung des Computereinsatzes in der Praxis des Arztes ist unbestritten. Es hat sich aber gezeigt, daß z.T. erhebliche Schwierigkeiten zu überwinden sind. So sind Tagungen gut geeignet, den augenblicklichen Stand der Entwicklung sowohl in der Planung, praktischen Anwendung oder in der Technik aufzuzeigen. Die vorliegenden Referate ermöglichen es nun auch den interessierten Nichtteilnehmern sich eingehend zu informieren.

Ge.

SHANNON, R. H. (Ed.)

Hospital Information Systems

Proceedings of the TC-4 Working Conference

on Hospital Information System,

Cape Town, South Africa, 2–7 April 1979

1979, 405 S., US \$ 58.50

North-Holland Publishing Comp., Amsterdam u. New York

Der Tagungsband enthält die während des Workshops gehaltenen Referate sowie übersichtsartig die in den Diskussionen angesprochenen Probleme. Dadurch entstand ein Überblick über den derzeitigen Stand der Entwicklung von Krankenhausinformationssystemen. In den Diskussionen wurden getrennt verschiedene Aspekte der jeweiligen Probleme angesprochen.

Die vorliegende Veröffentlichung sollte alle ansprechen, die sich über die Probleme im Zusammenhang mit der Verwirklichung und auch der Zukunft von Krankenhausinformationssystemen informieren wollen.

Ge.

LINDBERG, D. A. B.; REICHERTZ, P. L. (Ed.) Vol. 6

Lecture Notes in Medical Informatics

Health Care Technology Evaluation

Proceedings, Columbia, Missouri, Nov. 6–7, 1978

Editor: Goldman, J.

1979. 10 figs., 12 tab. VII, 118 pages,

Soft cover DM 25,-; approx. US \$ 13.80

Berlin–Heidelberg–New York: Springer-Verlag/ISBN3-540-09561-6

5 Übersichtsreferate, gehalten auf einem Symposium der Columbia-Universität, sind in vollem Wortlaut mit Tabellen, Grafiken und z.T. umfangreichem Schrifttumnachweis publiziert. Beigefügt ist die in weiten Teilen wörtliche Wiedergabe einer anschließenden Panel-Diskussion:

OMENN, G. S. u. BALL, J. R. »The Role of Health Technology Evaluation: A Policy Perspective«

FLAGLE, Ch. D. »An Overview of Evaluation Methods«

WILLEMS, J. S. »Cost-Effectiveness Analysis of medical Technologies as an Aid to Policymakers«

COLLEN, M. F. »Management of major Evaluation Efforts«

PAYNE, B. C. »Outcome Analysis in Health Care Technology Evaluation«

Kritische Selbstreflexion über Aufwand, Nutzen und Kosten, aber auch über politische Zielsetzungen, die mit solchen Großprojekten verbunden sind, zeigte die Notwendigkeit, die Methodologie der Kosten-Nutzen-Analyse der Volks- und Betriebswirtschaftslehre weiter und besser an die medizinische Technologie zu adaptieren.

Für Gesundheitspolitiker, die gewillt sind, sich an der Welt, wie sie ist, zu orientieren, eine Fundgrube für Erfahrungen, die nicht notwendigerweise ein zweites Mal gemacht werden müssen.

Fi.

WEBER, K.

BASIC für Anfänger

1979, 153 S., DM 31,-

Verlag Paul Haupt, Bern und Stuttgart

Die Bedeutung von BASIC als Programmiersprache gewinnt zweifellos in der letzten Zeit an Bedeutung. Daher ist es zu begrüßen, daß eine weitere systematische und übersichtliche Einführung vorliegt. Der Verfasser bemüht sich dabei, den Leser dieser Einführung dahin zu beein-

flussen, daß er nur »Standardanweisungen« benutzt, um portable Programme zu erstellen. Die jeweils eingefügten Beispiele erleichtern die Benutzung dieses Buches sehr.

Ge.

HEILMANN, H. (Hrsg.)

8. Jahrbuch der EDV (1979)

– Computerunterstützte Systementwicklung und Systemdokumentation

1979, 315 S., DM 79,-

Forkel-Verlag, Stuttgart – Wiesbaden

In den jährlich herausgegebenen Jahrbüchern werden immer ausgewählte Themenbereiche behandelt. Im 1979 Jahrbuch wird die Unterstützung durch die DV bei der Planung und Implementation von neuen Anwendungen der DV dargestellt. Die aufgezeigten realistischen Lösungen geben einen guten Einblick in den Stand der Entwicklung. – Dabei muß grundsätzlich unterstrichen werden, daß die DV Abteilungen für ihre eigenen Aufgaben sich des Hilfsmittels DV noch viel zu wenig bedienen. Daher sollte das vorliegende Buch ein nützliches Hilfsmittel für viele Praktiker der Softwareentwicklung sein.

Ge.

HEILMANN, H. u. W.

Strukturierte Systemplanung und Systementwicklung

1979, 302 S., DM 98,-

mit 88 S. Anhang

Forkel-Verlag, Stuttgart-Wiesbaden

Wenn wir z.Zt. eine Krise der EDV erleben, so ist es eine Krise der Softwareentwicklung. Die Kosten für die Software überschreiten heute schon vielfach die Kosten der Hardware. Dennoch will man es kaum hören, daß die Konzipierung, Entwicklung und Wartung der Software im Spannungsfeld zwischen Anwendern und EDV-Spezialisten nicht optimal ist. Hier versuchen die Verfasser mit der Darstellung von in der Praxis erprobten Methoden zu helfen. Hoffentlich wird die Hilfe auch angenommen.

Ge.

Bundesdatenschutzgesetz (BDSG)

Ergänzbarer Kommentar nebst einschlägigen Rechtsvorschriften

von H.-J. SCHAFFLAND und N. WILTFANG

Ergänzbare Ausgabe, einschließlich 7. Lieferung 750 S., DM 48,- zuzüglich Ordner DM 10,80

Erich Schmidt Verlag, Berlin-Bielefeld-München

Der Datenschutz personenbezogener Daten spielt eine immer bedeutendere Rolle. Hier ist es natürlich, alle relevanten Gesetze und Verordnungen des Bundes und der Länder zusammen mit Kommentaren und Zitaten der angesprochenen Bestimmungen aus anderen Gesetzesbereichen verfügbar zu haben. Die vorliegende Loseblattsammlung erfüllt diese Bedingungen und sollte daher heute in keiner EDV-Abteilung fehlen.

Ge.

Krebsatlas der Bundesrepublik Deutschland

Cancer Atlas of the Federal Republic of Germany

Krebssterblichkeit in den Ländern der Bundesrepublik Deutschland 1955–1975

Cancer Mortality in the States of the Federal Republic of Germany 1955–1975

Von/By Frentzel-Beyme, R., LEUTNER, R., WAGNER, G., WIEBEL, H. 1979, 45 farb. Karten, 15 Abb., 8 Tab. VIII, 70 Seiten (15 Seiten in Englisch).

Gebunden DM 88,-; approx. US \$ 48.40

Berlin–Heidelberg–New York: Springer-Verlag/ISBN3-540-09566-7

Das Deutsche Krebsforschungszentrum legt erstmalig für die Länder der Bundesrepublik Deutschland und für Westberlin zu den 10 epidemiologisch wichtigsten Organkrebsen Inzidenz- und Mortalitätsraten in ihrer zeitlichen Entwicklung zwischen 1955 und 1975 vor. Dies geschieht, wie in der deskriptiven Statistik üblich, in Form von Tabellen, Kurven und Kartografiken. Deren Auswahl und Gestaltung aber zeigt die Meisterschaft der Autoren, gleichermaßen dem statistischen Fachmann wie auch dem weniger Erfahrenen verdichtetes und standardisiertes Datenmaterial zu präsentieren.

Dieser Atlas ist unentbehrlich für alle auf den Gebieten Krebsforschung, -verhütung und -therapie Tätigen.

Fi.

RIEMANN, W. O.
Datenstrukturen – ihre Darstellung und Umsetzung
 1980, 78 S., DM 22,-
 Verlagsges. R. Müller, Köln

Bei der Datenverarbeitung, insbesondere beim Einsatz von Datenbanken, interessieren neben den Verarbeitungsalgorithmen hauptsächlich die zu bearbeitenden Datenstrukturen. In der vorliegenden Veröffentlichung werden u. a. die Klassifikation und die schematische Darstellung von Datenstrukturen sowie die weitgehend formalisierte Umsetzung in Speicherstrukturen behandelt.

Die klare Untergliederung in Datenstrukturen, Speicherungsstrukturen und die Realisierung der Datenstrukturen ergibt eine sehr übersichtliche und verständliche Einführung in eine nicht immer einfache Materie. Ge.

BEUTEL, P., KÜFNER, H. und SCHUBÖ, W.
SPSS 8 – Statistik-Programm-System für die Sozialwissenschaften
 3. Aufl.
 1980, 300 S., DM 22,-
 G. Fischer Verlag, Stuttgart – New York

Die vorliegende 3. Auflage berücksichtigt die Änderungen und Erweiterungen, die die Version 8 gegenüber der Version 7 gebracht hat. Dabei wurden aber die notwendigen Textänderungen so eingefügt, daß diese Ausgabe auch noch für die Versionen 6 und 7 benutzbar ist.

Bei der Verbreitung des Programm-Systems SPSS ist die vorliegende deutsche Bearbeitung für die Benutzer ein wirklich nützliches Hilfsmittel. Ge.

MERTEN, K. und LISCH, R.
FORTRAN IV
 2. überarbeitete Aufl., 1980, 210 S., DM 16,80
 UTB Taschenbuch Band 428
 G. Fischer Verlag, Stuttgart – New York

Es gibt eine Fülle von Einführungen in die Programmierung und so auch in die Benutzung der Programmiersprache FORTRAN, dennoch sollte man das vorliegende Taschenbuch nicht übersehen. Die jetzt vorliegende überarbeitete 2. Auflage kann vorbehaltlos speziell Anfängern empfohlen werden, weil immer wieder versucht wird, die Zusammenhänge zwischen den verschiedenen Möglichkeiten deutlich zu machen. Die dargestellten Beispiele und Übungen sind überschaubar und doch recht instruktiv. Ge.

WEBER, K.
BASIC – Programmierrichtlinien für Austauschprogramme
 1979, 81 S., DM 33,-
 Verlag Paul Haupt, Bern und Stuttgart

Zusammen mit den Mini- bzw. Mikrocomputern wird BASIC als Programmiersprache immer mehr benutzt. Es ist daher naheliegend, daß man sich Gedanken über eine Vereinheitlichung dieser Programmiersprache macht, um erarbeitete Programme austauschen zu können. Es ist daher zu begrüßen, daß in dem vorliegenden Buche verschiedene »Dialekte« tabellarisch gegenübergestellt wurden und aufgrund der in Arbeit befindlichen Normen Richtlinien aufgestellt wurden, um transportable Programme zu erstellen. Für alle BASIC Programmierer ein nützliches und notwendiges Hilfsmittel. Ge.

BARTENWERFER, A. und RAATZ, U.
Einführung in die Psychologie Bd. 6
 -Methoden der Psychologie
 1979, 185 S., DM 16,80
 Akademische Verlagsges. Wiesbaden

Die Autoren grenzen im Vorwort selbst den dargestellten Bereich ein, wenn sie schreiben, daß mehr die Methoden bei der Erarbeitung von Forschungsergebnissen behandelt werden sollen. So werden neben allgemeinen Fragen in den ersten Kapiteln (»Psychologie und Wissenschaft« und »Fragen allgemeinen methodischen Vorgehens bei erfahrungswissenschaftlicher Forschung in der Psychologie«) dann »Spezielle Fragen bei der Planung einer Untersuchung«, »Methoden der Datenerhebung« und »Methoden der Datenauswertung« angespro-

chen. Der zur Verfügung stehende Platz (insgesamt nur 185 Seiten) bedingt, daß viele Dinge nur stichwortartig, stellenweise sogar nur durch einen Literaturhinweis dargestellt werden konnten. So bleibt insgesamt eine sehr knappe Einführung, die aber als Repetitorium gute Dienste zu leisten vermag. Ge.

GRIESSER, G.G. et al. (Ed.)
Data Protection in Health Information Systems
 Considerations and Guidelines
 1980, 218 S., US \$ 44.00
 North-Holland Publishing Comp., Amsterdam and New York

Das vorliegende Buch richtet sich insbesondere an alle, die sich ohne nennenswerte EDV-Kenntnisse mit der Planung oder dem Einsatz von Krankenhausinformationssystemen beschäftigen wollen. Es entstand aus einer IMIA-Arbeitsgruppe und ist mit der klaren Gliederung stets als Leitfaden für den Einsatz von Krankenhausinformationssystemen geeignet. Ge.

WOERZ-BUSEKROS, A.
Algebras in Genetics
 Lecture Notes in Biomathematics Vol. 36
 1980, 237 S., DM 28,50
 Springer-Verlag, Berlin – Heidelberg – New York

In der vorliegenden Abhandlung ist eine möglichst vollständige Darstellung der mathematischen Theorie der Algebra in der Genetik versucht worden. Dabei werden eine Fülle von Anwendungen auf konkrete genetische Situationen betrachtet. Ein umfangreiches Literaturverzeichnis ergänzt die übersichtliche Darstellung. Ge.

AN DER HEIDEN, U.
Analysis of Neural Networks
 Lecture Notes in Biomathematics Vol. 35
 1980, 159 S., DM 23,50
 Springer-Verlag, Berlin – Heidelberg – New York

Im Mittelpunkt dieser Abhandlung steht die mathematische Theorie; auf mögliche Anwendungen wird z.T. nur durch Literaturangaben hingewiesen. Die Netzwerke werden durch Systeme von nichtlinearen Integralgleichungen beschrieben. Daher kann die vorliegende Darstellung auch als Kurs in nichtlinearer Systemtheorie verstanden werden. Dies gilt umso mehr als die mitgeteilten mathematischen Ergebnisse auch auf andere Bereiche übertragen werden können. Ge.

JANKO, W.H.
APL I
 Eine Einführung in die Elemente der Sprache und des Systems
 1979, 183 S., DM 14,80
 Athenäum Taschenbuch, AT 5023
 Athenäum Verlag, Königstein/Ts.

Dieses Buch wendet sich an jene, die APL in einer einführenden Version kennenlernen wollen. Obwohl es nur eine Teilmenge des Sprachumfangs und des Umfangs an Systembefehlen beinhaltet, kann der gebotene Stoff insofern als abgeschlossen angesehen werden, als es dem Leser nach der Lektüre dieses Buches möglich sein sollte, einfache Programme in APL zu schreiben. Für jene, die APL intensiver nutzen wollen, sei auf den Fortsetzungsband hingewiesen. Ge.

WALTER, E.:
Biomathematik für Mediziner
 2. Aufl., 1980, 206 S., DM 19,80
 Teubner Studienbücher
 B. G. Teubner, Stuttgart

Die zweite Auflage dieser didaktisch sehr gut geschriebenen »Biomathematik für Mediziner« ist dem geänderten Gegenstandskatalog für

die ärztliche Prüfung angepaßt worden. Der Bezug zum Gegenstandskatalog wurde bei jedem Abschnitt hergestellt; zusätzliche Abschnitte wurden besonders gekennzeichnet. Die Darstellung ist knapp und klar und durch verständliche und instruktive Beispiele erläutert. Ein empfehlenswertes Buch für Medizinstudenten, aber auch für viele Praktiker, die sich schnell orientieren und informieren wollen. Ge.

DEMMIG, G.:

Programmieren – leicht gemacht

Teil 1 und Teil 2

Teil 1, 1977, 96 S., DM 13,80

Teil 2, 1978, 121 S., DM 13,80

Demmig Verlag KG., Bad Nauheim

Es fehlt im Augenblick wirklich nicht an Versuchen, leicht verständliche Einführungen in die Programmierung zu schreiben. Bei dem schnellen Anstieg der Kleincomputerinstallationen wächst dabei auch die Bedeutung der Programmiersprache BASIC.

In den vorliegenden Büchern wird versucht, eine Einführung in BASIC zu geben und gleichzeitig die Unterschiede zu FORTRAN und ALGOL aufzuzeigen. Die gewählten Beispiele sind aus den unterschiedlichsten Anwendungsbereichen genommen, wobei die notwendigen Formeln zum Teil ausführlich erläutert werden. Ge.

Lecture Notes in Biomathematics

Managing Editor: Levin, S.

Vol. 34

Collins, R., van der Werff, T. J.

Mathematical Models of the Dynamics of the Human Eye 1980. 25 figs., 3 tabb. VII, 99 pages.

Soft cover DM 19,50; approx. US \$ 11.50

Berlin – Heidelberg – New York: Springer-Verlag

Diese Monografie will eine Zusammenschau von 257 Originalarbeiten geben, welche Beziehungen zwischen physiologischen Meßgrößen des Auges untereinander oder solcher mit Meßgrößen des übrigen Organismus funktional zu beschreiben versuchen. Hierbei beschränken sich die Autoren fast ausschließlich auf sog. kausale Modelle. Dies sind entweder solche, die sich unmittelbar aus der physikalischen Strömungslehre herleiten, oder aber andere, die – entsprechend den Gepflogenheiten einer physikalisch orientierten Physiologie – das Geschehen im Organismus durch einen Satz einfacher oder partieller Differentialgleichungen beschreiben.

Beide Wege führen indes auf sehr komplizierte Exponentialfunktio-

nen. Die Schätzung der »Konstanten«, besser Parameter dieser Funktionen durch statistische Methoden erbringt in der Praxis aus vielerlei Gründen fast immer instabile Lösungen. Es fragt sich, ob man weniger anspruchsvoll mit formalen Modellen nicht weiterkäme.

Unberührt hiervon bleibt jedoch das Verdienst der Autoren dieser Monografie. Denn dieses liegt mehr in der Übersicht über den gegenwärtigen Wissensstand, in der Erschließung der Originalliteratur durch Schlagwörter, in der Definition medizinischer Begriffe, mit denen die Biomathematiker ihren Modellbau betreiben, sowie in dem Versuch, in das heillose Durcheinander der Notationen physiologischer Konstanten, Parameter sowie Einfluß- und Zielgrößen Ordnung zu bringen. Experimentelle und klinische Ophthalmologen, Physiologen sowie Biomathematiker erhalten eine schnelle und präzise Einsicht. (Fi.)

Medizinische Informatik und Statistik

Herausgeber: KOLLER, S.; REICHERTZ, P. L.; ÜBERLA, K.

Band 18

Fischer, R.-J.

Automatische Schreibfehlerkorrektur in Texten

Anwendung auf ein medizinisches Lexikon

1980. 5 Abb., 8 Tabb. X, 89 Seiten. Geh. DM 25,-;

approx. US 14.80

Berlin – Heidelberg – New York: Springer-Verlag

Im beginnenden Zeitalter der EDV-gesteuerten Schreibautomaten erwartet der Anwender im medizinischen Bereich, daß diese auch durch Personen bedient werden können, deren Kenntnisse in der medizinischen Fachnomenklatur eher bescheiden sind. Dies führt zu der Forderung, daß die Korrektur von Schreibfehlern medizinischer Termini automatisch aus einem gespeicherten Wortvorrat erfolgen soll.

Das vorliegende Buch gibt eine Übersicht der hierfür geeigneten und diskutierten Algorithmen, stellt einen Katalog von Anforderungen an solche Verfahren auf, untersucht mengentheoretisch deren Begründungen und Folgerungen und erläutert schließlich das HAMMING-Prinzip ausführlich.

Ein im Anhang aufgelistetes PL-1-Quellenprogramm wird in seiner Struktur und Philosophie beschrieben. Die empirischen Ergebnisse der Anwendung auf eine Zufallsauswahl von 1000 medizinischen Fachbegriffen aus dem AGK-Thesaurus, der 33 952 Wörter umfaßt, werden mitgeteilt, wobei interessante deskriptive Statistiken der medizinischen Fachsprache anfallen.

Allen, die an automatischer Textkorrektur arbeiten, sollte dieses Buch Pflichtlektüre sein. Darüber hinaus sollte es jeder medizinische Informatiker gelesen haben, wird er doch mit Sicherheit irgendwann mit solchen Problemen konfrontiert. (Fi.)